AUTOMATIC CONTROL

PRINCIPLES and PRACTICE

HOLZBOCK

AUTOMATIC CONTROL LIBRARY

AUTOMATIC CONTROL

PRINCIPLES AND PRACTICE

WERNER G. HOLZBOCK
Chief Engineer
Askania Regulator Company
Chicago, Illinois

REINHOLD PUBLISHING CORPORATION
NEW YORK
CHAPMAN & HALL, LTD., LONDON

Library of Congress Catalog Card Number 58-9742

REINHOLD PUBLISHING CORPORATION

Publishers of Chemical Engineering Catalog, Chemical Materials Catalog, "Automatic Control," "Materials in Design Engineering," "Progressive Architecture"; Advertising Management of the American Chemical Society

Printed in U.S.A. by
THE GUINN CO., INC.
New York 1, N. Y.

PREFACE

This book tries to answer the need for a comprehensive text on automatic control without entering into differential equations.

After clarifying some basic concepts and terms in the first chapter, static characteristics are discussed in Chapter 2. It is hoped that sufficient emphasis is given to static prerequisites which are frequently neglected in concentrating on the dynamic aspects. This does not impair the importance of dynamics and this subject is extensively treated in Chapters 3 to 5.

After this introduction to operational behavior, the hardware of control systems is dealt with in Chapters 6 through 8. It is felt that the first five chapters will prepare for the discussion on hardware and provide a basis for the understanding of its decisive role.

Chapters 9 through 11 describe the elements of the control loop: the measuring element, the controller itself, and the final control element. Particular emphasis has been put on temperature control and control valves, since both are elements whose behavior frequently limits the potential benefit of automatic control.

The remaining two chapters are concerned with the use of controllers in the system. Various controls—averaging, ratio, cascade, etc.—are discussed and a number of systems from various industries are described as examples.

This text is within the grasp of the intelligent technician. This does not mean that it can be read without effort. The nature of the subject material does not allow a facile text which can be digested by the reader without exertion on his part.

The student interested in control theory will find this book a primer which prepares him for more profound treatment of the subject in a number of more theoretical books of excellent quality.

The engineer who wants to get a practical understanding of controllers and control systems will appreciate this text which foregoes the refinements of theoretical analysis and stresses physical realities.

The method of presentation which I have used in this book is the

result of many talks and publications. I have published articles in *Air Conditioning, Heating and Ventilating, Chemical Engineering, Control Engineering, Regelungstechnik* and other leading journals, and many a thought which is presented in the following pages was first discussed in these articles, as well as in talks to ISA and ASME groups.

WERNER G. HOLZBOCK

Evanston, Illinois
March 1958

CONTENTS

1. THE AUTOMATIC CONTROL SYSTEM

With few exceptions, and these will be pointed out, this text follows the terminologies prepared by the American Society of Mechanical Engineers and published in their Standards 105 and 107. Definitions that are extracted from these standards are made with the permission of the publisher.

The Control Loop

The elements of an automatic control system are combined in a closed loop, as illustrated in Figure 1–1. The measuring means, automatic

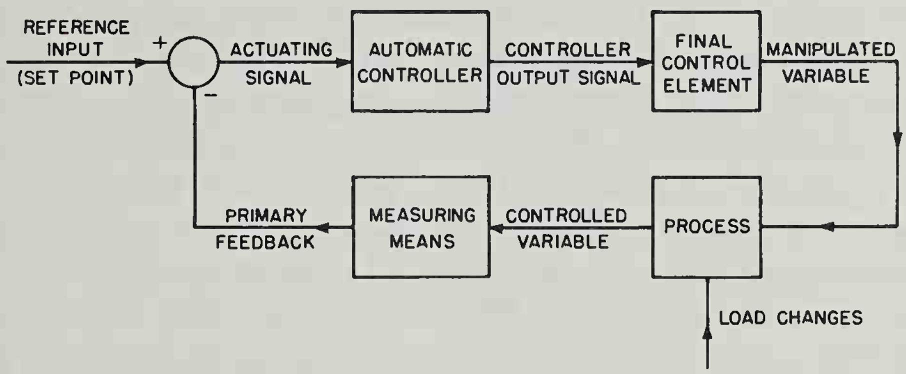


Figure 1–1. Automatic control system.

controller, and final control element are considered separate units. This is in agreement with common usage but differs from the ASME terminology, which combines all three units under the term automatic controller and leaves no term to designate the unit which is generally known as the controller proper.

The input to the automatic controller is the actuating signal which is the difference at any time between the reference input and the primary feedback. This difference is obtained in a component called the summing point which is symbolized by the circle in the illustration. The summing point is generally part of the controller.

The primary feedback is the magnitude of the controlled variable as interpreted by the measuring means. Hence, the actuating signal can also be defined as the deviation of the controlled variable from the set point.

The automatic control system includes the process. In a more specific sense, however, the control system may be considered as the combination of measuring means, controller, and final control element, disregarding the process. Thus, in speaking of the hardware of a control system, process elements like cracking towers, evaporators, etc., are usually excluded.

The purpose of the automatic controller is to reduce the actuating signal to a minimum, and preferably to eliminate it, with the least delay possible. The means of effecting this is the controller output signal which is transmitted to a final control element. This may be a valve, a pump or any other device which directly changes the value of the manipulated variable. The manipulated variable is a condition or characteristic of the control agent. For example, when a final control element changes the fuel-gas flow to a burner, the manipulated variable is flow, and the control agent is fuel gas.

The controlled variable is a condition or characteristic of the controlled medium. For example, where temperature of water in a tank is automatically controlled, the controlled variable is temperature and the controlled medium, such as heat, is utilized or transformed by the primary element to produce an effect which is a function of change in the value of the controlled variable. The effect produced by the primary element may be a change of pressure, force, position, electrical potential, or resistance. The primary element is a portion of the measuring means which consists of those elements involved in ascertaining and communicating to the summing point of the controller the value of the controlled variable. The signal which the measuring means transmits to the summing point is the primary feedback. It is compared with the reference input signal which in turn is determined by the set point.

The set point is the position to which the control-point setting mechanism is fixed. Where the automatic controller possesses a set-point scale, the set point is the scale reading translated into units of the controlled variable. Where a setting scale is not provided, the set point is the position of the control-point setting mechanism translated into units of the controlled variable. The set point may be varied manually or automatically, as in time-schedule or ratio control systems.

In some types of automatic controllers—for example, those with two-position differential-gap action or floating action with neutral zone—the set point is related to the position of a range of values of the controlled variable. The set point is often selected as the center of this range of values.

A permanent deviation of the controlled variable from the set point is called offset. It is convenient to speak of the value for which the

controller is adjusted as set point and of the value which the controller maintains as control point. The control point is the set point plus or minus offset.

The disturbances, called load changes, entering the process represent the variable that requires adjustments by the controller. Without load changes the control system could just as well be manual, since once it is set no changes will take place.

Load changes have various causes. Figure 1–2 shows a reboiler into which the process fluid enters as condensate and is brought to a boil by

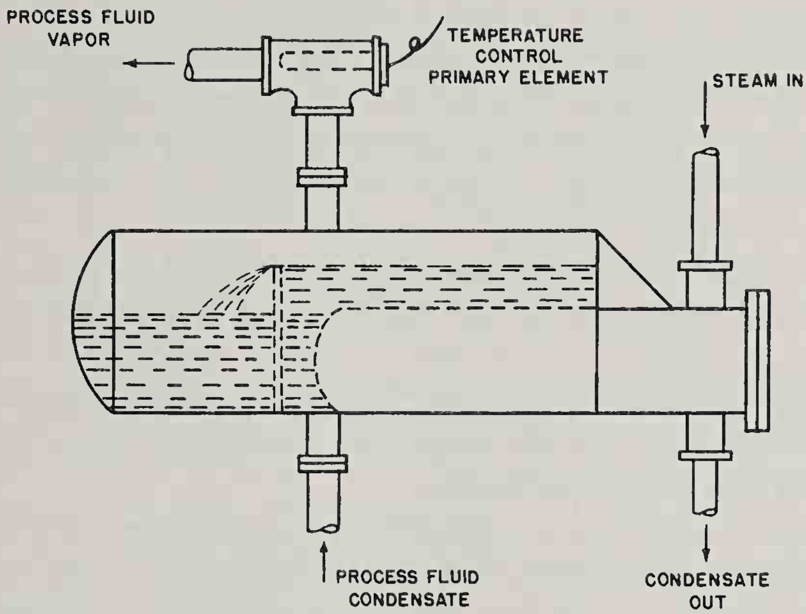


Figure 1–2. Reboiler.

a steam heat exchanger. The process fluid leaves the reboiler as vapor. A level controller (not indicated) maintains the level. The vapor temperature is the controlled variable and steam flow is the manipulated variable. The load changes of this process are affected by (a) temperature of entering process fluid, (b) steam temperature, (c) steam pressure as reflected in rate of flow change for a given valve opening, and in condensation temperature, and (d) outside temperature, draft cooling, etc. In addition to these load changes, there are other long-range but not less important changes like scaling of tubes, deterioration of heat insulation, etc.

Time Lags

If actions and reactions in a control system were to take place without loss of time, no control problem would exist. For example, the steam production of a boiler operating under such ideal conditions would

increase upon the slightest drop in steam pressure and balanced conditions would be restored immediately and without difficulty.

These balanced conditions require that the input of steam-producing elements correspond to the steam output as required by the load. Practical circumstances produce a time interval between the output change, the necessary corrections, and the taking effect of these corrections.

The first delay occurs due to the fact that a load change is not sensed immediately, but after it results in a change of the controlled variable, i.e., steam pressure in the above example. This means that the control system which is intended to keep the controlled variable constant is built on the premise that the controlled variable changes before corrective action takes place. This is an unavoidable contradiction inherent in all automatic control systems. The art is to keep the change to a minimum. If, for example, a controlled variable is to be kept within ± 1 per cent of its set point value, it is necessary that a much smaller change than this suffices to initiate corrective action.

The next delay occurs in converting the controller output signal into a change of the magnitude of the manipulated variable by means of the final control element.

Frequently, the most serious delay is due to the time lost in the process before the controlled variable responds. Two main causes produce this delay: resistances and capacitances.

Resistance is opposition to flow. It is expressed in units of potential change required to produce a unit change in flow rate. For example, the flow of heat from the steam to the water through the tubes of a heat exchanger is delayed because of the thermal resistance caused by scales, gaseous films, etc.

Thermal resistance is the temperature differential in degrees Fahrenheit* required to produce a heat flow rate of one Btu (British thermal unit) per hour through a given body.

Fluid resistance is the psi fluid pressure differential required to produce a fluid flow rate of 1 cu ft/min. through a given passage.

Electrical resistance is the voltage necessary to produce a current of one ampere (one coulomb per second) through a conductor.

Capacitance is the change in quantity contained per unit of change in a reference variable; for example, the speed with which the liquid level in a tank changes depends on the change of volume of stored liquid per unit change of head. If a tank level drops one foot when 20 cu ft of liquid are removed, then its capacitance is 20 cu ft/ft or 20 sq ft which

---

*Units in the following definitions for resistances and capacitances are those most commonly used in practice.

is equivalent to the area of the liquid surface. If the shape of the tank causes the liquid surface area to vary with change of head, the capacitance will likewise vary with head.

In controlling the pressure in a gas-filled tank, the capacitance may be expressed as a function of the weight of gas. In this case the capacitance is the change of weight of stored gas per unit change of pressure.

The most typical capacitance is probably the flywheel. All capacitances act like flywheels which are storing up energy, reducing rapid response to a change.

Thermal capacitance is the Btu's absorbed by a body per degree Fahrenheit rise in its temperature.

Volume capacitance is the cubic feet of solids or liquids that can be stored in a container per foot of increase in level. In the case of gases it is standard cubic feet*/psi of pressure change.

Weight capacitance is the pounds of solids or liquids that can be stored in a container per foot of increase in level. In the case of gases it is standard cubic feet under standard conditions/psi of pressure change.

Electrical capacitance is the change of electrical charge of a capacitor expressed in microfarad/volt across its terminals.

Capacitances produce a delay which is expressed in the time constant and which will be discussed in greater detail further below. In addition to the time constant, a delay exists between two related actions in a control loop; this is called the *dead time.*

Figure 1–3 illustrates a pH control system. The control valve regulates a chemical which maintains the pH of the main flow. Before corrective action of the control valve reaches the pH measuring means,

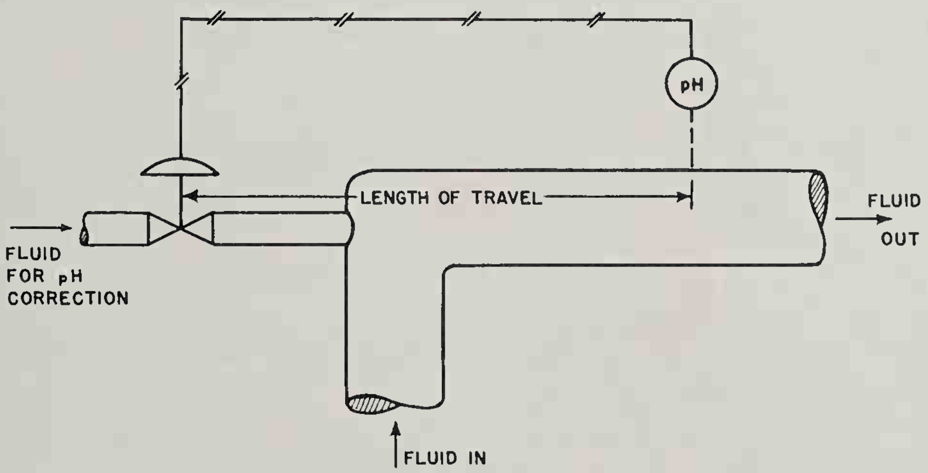


Figure 1–3. pH control system.

*Standard cubic feet refers to the volume of the gas at a pressure of 29.921 in. Hg (760 mm Hg) and a temperature of 59°F (15°C).

some finite time elapses which is a function of the length of travel. The time thus lost is dead time and plays an important part in the stability of the system, as will be shown. It is paramount in any control system that dead time be kept as low as possible.

Speed of Corrective Action

If the corrective action in a system with time lags would be just enough to bring the system back to its control point after a load change, the action would be too slow, the controlled variable would continue to change in the meantime, and close control would be impossible in the majority of cases.

On the other hand, intensifying the corrective action after a load change—for example, admitting more fuel to a process than is actually necessary, in order to gain faster action—produces overshooting. When the controlled variable, e.g., pressure, is back at its control point, the excessive fuel admission results in overcorrection. This changes the pressure in a direction opposite to its initial deviation, and the resulting control action will now produce undershooting. Oscillations around the control point are the result.

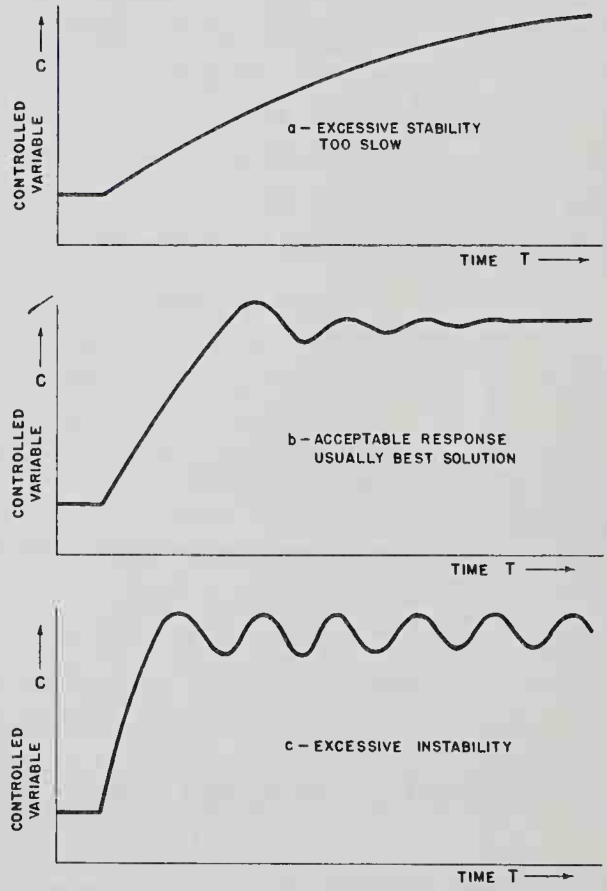


Figure 1–4. Various process responses after a change in set point.

The entire problem of automatic control centers around maintaining a control system within acceptable stability conditions while obtaining the greatest possible speed in corrective action.

Stability

Figure 1–4 shows various forms of responses of a controlled variable to a sudden change in set point of the controller. The controlled variable may be pressure, level head, temperature, etc. The shape of these response curves expresses the stability of the system. In curve *a*, the new value is approached without overshoot; however, considerable time elapses until reaching the new value. In curve *b*, the response is considerably faster. There are a few rapidly subsiding oscillations. They are usually acceptable and this form of response is generally considered the most desirable. Curve *c* shows a response of almost continuous oscillations. In general, such a response is not acceptable. Occasionally, however, such oscillations, provided they are within limited amplitudes, can be tolerated. In such cases discontinuous controllers, described in a later chapter, would probably give a satisfactory and economical solution.

2. STATIC CHARACTERISTICS

Accuracy, resolution sensitivity, dead band, hysteresis and linearity are static characteristics of the components of an automatic control system which determine its operational quality. Although these characteristics must meet the requirements of the case, unnecessary refinements must be avoided for reasons of economy. A thorough knowledge of these static characteristics is necessary for realistic appraisal.

One of the difficulties in discussing static characteristics is the overlapping of terms. Accuracy includes scale and calibration errors as well as the results of resolution sensitivity, dead band, hysteresis, and linearity. Resolution sensitivity is due to static friction, frequently called break-loose force, or to discontinuities such as the windings of a potentiometer. Dead band includes static friction and discontinuities, as well as lost motion. Hysteresis is, strictly speaking, only electrical or mechanical, and is a material property that will be described later. However, since in control system components hysteresis can generally not be separated from dead band and nonlinearity, it may occasionally be convenient to consider hysteresis as including hysteresis proper, dead band, and nonlinearity.

Accuracy

Figure 2–1 is the block diagram of an active component in a control system. Input and output are related by a mathematical function which expresses the correct value.

Figure 2–1. Component in control system.

For example, the component of Figure 2–1 may be a pneumatic controller which includes the control-point setting mechanism. The input may be a temperature, and the output, a pneumatic signal. At 60°F, the controller output signal is 3 psi, and at 120°F, it is 15 psi. The mathematical function that applies for the range from 60° to 120°F is

$$P = 0.2T - 9$$

where P is the correct value of the controller output signal and T is the input temperature.

The output value differs from the correct value because of inherent imperfections of the controller. The degree to which the output value approaches the correct value is the *accuracy*.

Accuracy may be expressed in units of the input, in per cent of the output range, in per cent of the value at which the accuracy was determined, etc. A per cent specification which is not related to some specific magnitude is meaningless.

In the case of the aforementioned temperature controller with an output of 3 to 15 psi, the accuracy may be stated as 1 per cent of range. If the controller has a range of 100°F, the output signal differs from the correct value by not more than ± 0.2 psi. This is derived by considering that 1 per cent of 100° is equal to 1°; since the output signal changes 12 psi for 60° change, the equivalent of 1 degree is 0.2 psi. This is the pressure by which the output may be above or below the correct value.

The specification of accuracy by a number leaves open several questions which relate mostly to the conditions under which the accuracy was determined. For example, temperature and relative humidity may have considerable influence on the accuracy. A pneumatic controller can show inaccuracies with changes in air temperature, which may not be reflected in an accuracy statement. Acceptable changes in supply voltage, line frequency, supply air pressure, and hydraulic pressures may have to be specified before an accuracy rating becomes meaningful.

A Bourdon tube pressure gauge, for example, usually is accurate within ± 0.5 per cent of its range. This gauge responds to the difference between process pressure and atmospheric pressure. In fact, it is an instrument that measures differential pressure. If it is calibrated for 15 psig (lb/sq in. gauge pressure), it really measures about 30 psia (lb/sq in. absolute pressure) minus the atmospheric pressure. Let this gauge be calibrated for an atmospheric pressure of 15 psia. Since this pressure may vary between ± 5 per cent, it follows that the gauge pressure has an accuracy of ± 2.5 per cent of its range due to atmospheric pressure conditions alone. To this must be added a temperature error which is usually in the order of plus 2 per cent for a temperature rise of 100°F.

Accuracy specifications are further left open if the accuracy figure is the result of a number of readings of which it is the average, and if it refers to any value of the input or is only attainable for certain values. In order to give a fair indication of accuracy, the worst reading over the whole range should be the one reported. If this is not the case, then the conditions to which the accuracy refers should be stated. This, of course, is equally true of other specifications like dead band, resolution sensitivity, etc.

Calibration Errors

Calibration errors belong to one or both of two categories: zero errors and angular errors. The zero error is a linear shifting of the range. For example, the output may read two units too high at any point throughout its range. This is a zero error. The adjustment consists in linearly shifting the range to obtain correct zero reading. The angular error shows correct reading at an arbitrary point of the scale, e.g., the zero point, and an error which increases in proportion to the distance from this point. This requires adjusting the ratio of the output-input relationship which is usually accomplished by changing a lever ratio in a mechanical linkage or the gain in an electric circuit to increase or decrease the output signal range for a given input range. Zero errors and angular errors are frequently combined in the overall calibration error, and require separate adjustments.

Resolution Sensitivity

Resolution sensitivity is the minimum change in the measured variable which produces an effective response of the component in a control system. Suppose the resolution sensitivity of a spring-opposed diaphragm control valve is to be determined. The input is an air-pressure signal. The output is stem motion. The air pressure is raised to an arbitrary value and the stem is allowed to come to rest. Next, the air pressure is further increased, but now by very small increments in order to determine the minimum increment required to start the stem moving. Let this minimum increment be 0.012 psi. For a pressure range of 3 to 15 psi, this means a resolution sensitivity of 0.1 per cent.

Resolution sensitivity in this case expresses the force that is required to induce a physical system to produce motion. It is due to static friction, sometimes called stiction.

Another cause of limited resolution sensitivity is due to a built-in stepping arrangement, such as a wire wound rheostat. The resistance can be changed only in discrete steps since one winding after another is shunted out and each winding is of some finite resistance. The resistance in one winding corresponds thus to the resolution sensitivity of the rheostat.

In the case of the rheostat, static friction is probably also present and the resolution sensitivity is caused by both static friction and winding resistance. In no case is lost motion included in resolution sensitivity as it is in dead band. Hence from the viewpoint of actual operating conditions, the significance of resolution sensitivity becomes questionable. The input signal may either increase or decrease and the essential ques-

tion is how much the input can change without producing a response in the output. However, this is dead band and not resolution sensitivity. In the test for the latter, the signal is changed in one direction only and is not reversed as in actual operation. The specification of resolution sensitivity is of interest only if it is desirable to separate lost motion in the analysis.

Dead Band

The range of values through which the input can be varied without initiating output response is defined as dead band. This is not only the limiting factor for accuracy in a component, but is also the most important operational quality criterion of static characteristics for an automatic control system. If a flow controller with a range of 100 gpm has a dead band of ± 1 per cent of range, this means that if it is to control at 60 gpm, the actual flow may vary between 59 and 61 gpm without producing action.

If the flow has been at 59 gpm and now begins to increase, no action may take place until the flow is 61 gpm. This means not only that the controlled variable is able to change through 2 gpm without correction taking place, but also that if the flow continues to increase, the controlling action is delayed. This is equivalent to dead time and contributes to the instability of the control loop, as will be discussed in the following chapter.

The cause of dead band is static friction and discontinuities, i.e., resolution sensitivity plus lost motion. The most common cause of lost motion is probably backlash, which refers to the looseness of mechanical linkages, gears, etc. There are, however, other causes of lost motion, such as the overlap between spool lands and ports in valves, as will be discussed in the chapter on Hydraulic Control Elements.

Linearity

Linearity can be expressed in a number of ways as shown in the following:

Normal or Independent Linearity. This is illustrated in Figure 2–2 and expresses the maximum deviation of the output values (full line) from the best straight line (dashed line) that can be drawn through these values. Considering a scale of ten units, the maximum deviation in Figure 2–2 is about one unit or ten per cent.

Zero-based Linearity. This is illustrated in Figure 2–3. The dashed line to which the linearity is referred passes through zero. In the illustration, the maximum deviation from this line is about 1.3 units or 13 per cent.

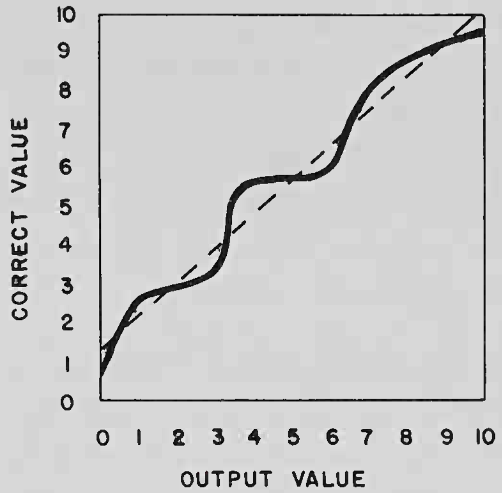


Figure 2–2. Normal linearity.

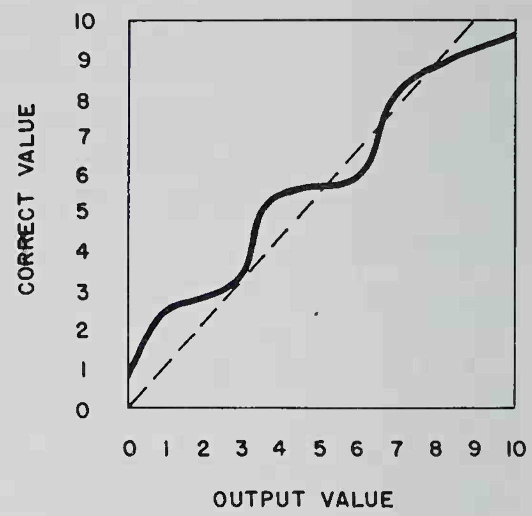


Figure 2–3. Zero-based linearity.

Absolute Linearity. Neither the normal nor the zero linearity expresses the accuracy of the output value as compared to the correct value. The resulting deviation from linearity is considerably larger and is shown in Figure 2–4. The dashed reference line is the idealized output which corresponds with the correct value. The greatest deviation is now about 1.6 units or 16 per cent.

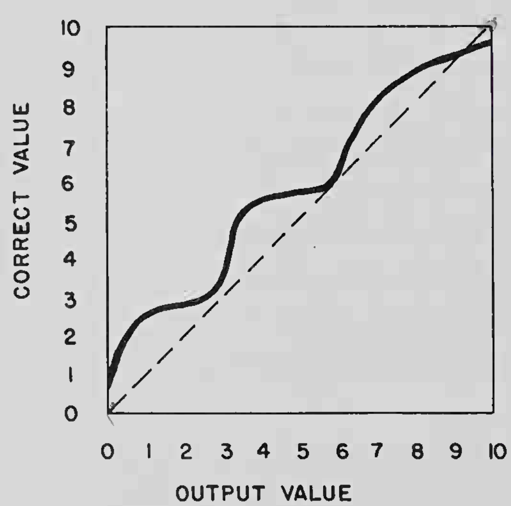


Figure 2–4. Absolute linearity.

In no case does linearity necessarily include backlash or hysteresis. It consists of values obtained by changing the input signal in one direction only. The causes of nonlinearity are general mechanical or electrical imperfections.

Hysteresis

It was stated before that the hysteresis curve of a control component generally can not be isolated from the effects of dead band and linearity,

and that it may therefore be convenient to consider a hysteresis curve as an agglomeration of these characteristics. Hysteresis by itself, however, is part of the inherent physical characteristics of materials used. It is distinguished by this from all other characteristics here discussed which are based on the interrelation of the elements in a component. (The

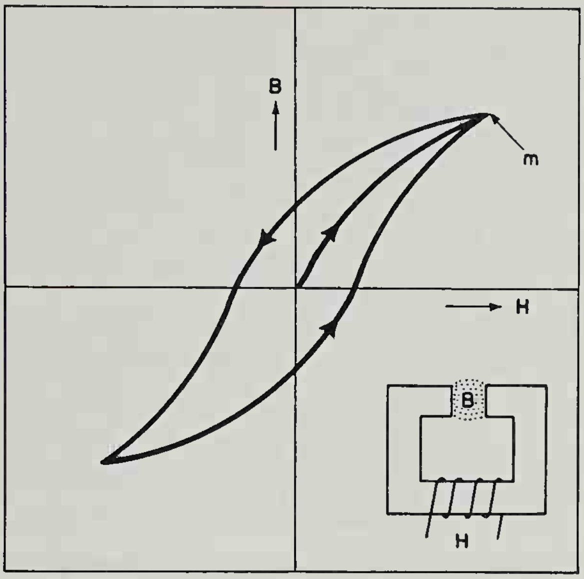


Figure 2–5. Magnetic hysteresis.

jumps in discrete quanta as known in physical science are ignored in this connection as being irrelevant in the practical behavior of control systems.)

Hysteresis may be magnetic or mechanical, as shown in Figures 2–5 and 2–6, respectively. Both graphs show output-input relations of a

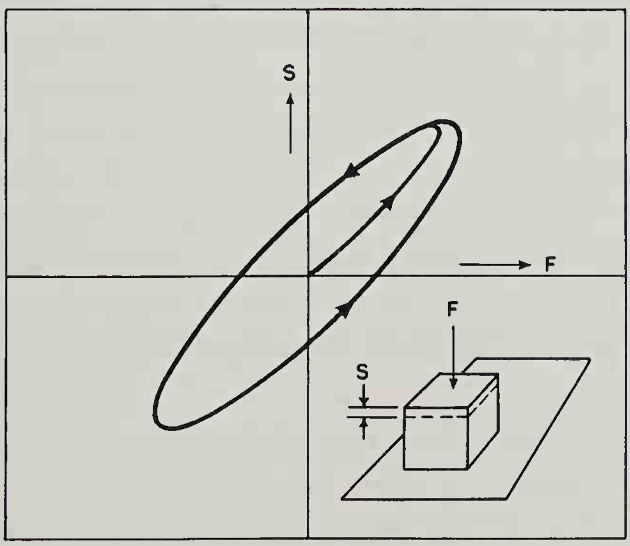


Figure 2–6. Mechanical hysteresis.

physical specimen. In the case of Figure 2–5, the input is magnetic intensity H produced by and proportional to the current in the winding. The output or result of the magnetic intensity is the magnetic flux B.

In Figure 2–6, the input is the stress F, a force applied to the elastic specimen, which results in a strain or deformation S, equivalent to the output.

In both cases, the output-input relation depends on the previous history of the specimen. Initially, they are both at zero, but once the input signal has been applied, there remains a memory of this signal in the specimen, called the retentivity or the remanence. The magnitude of this remanence depends on the amplitude of the input signal. If the input signal is cycled through smaller and smaller cycles, the hysteresis becomes less and less noticeable. It is therefore necessary to specify the amplitude of the input signal when giving hysteresis values.

Occasionally, dead band and hysteresis are used as equivalents. Dead band, however, refers to a change of input without output response, under which conditions hysteresis does not appear at all. The two terms relate to entirely different characteristics.

Linearity and Hysteresis in the Control Systems

Wherever a reading of the controlled variable is required, both linearity and hysteresis in the measuring means affect directly the accuracy of the reading and are important criteria. The controlling function, however, can tolerate considerable nonlinearity and hysteresis in any part of the loop without deterioration of performance. Figure 2–7 shows a simplifi-

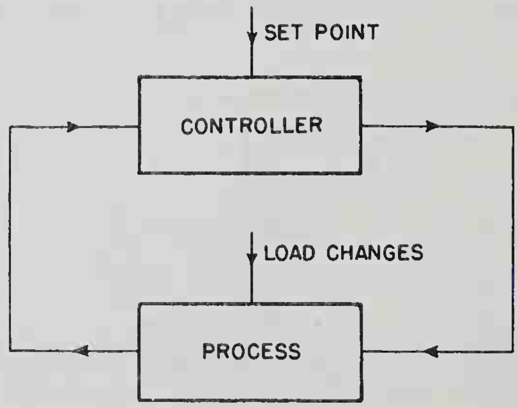


Figure 2–7. Basic control system.

cation of Figure 1–1 in which the control system is reduced to a controller and a process. In this case, the controller is supposed to incorporate both measuring means and final control element. The controller acts upon the process, and the process acts upon the controller. A

change of input from the process to the controller results in a modification of output from the controller. This produces a correction in the process and a resultant modification of the input to the controller coming from the process. If these steps were to be immediate ones and happen in the indicated succession only once, the linearity and hysteresis of the controller would be very important. The physical reality however, is that these changes occur gradually, though they may be fast, and because of the interactions in the closed loop, a self-correcting system is set up in which many nonlinearities can be absorbed.

A typical example is the industrial control valve which will be discussed in a separate chapter. Some pertinent facts are anticipated here. The control valve is part of the controller in Figure 2–7 and adjusts the manipulated variable, i.e. flow, which is part of the process. The relation between change of the manipulated variable and response of the controlled variable is a characteristic of the process and generally not subject to alteration. The relation is not a linear one and, furthermore, depends on the nature of load changes. In order to compensate at least partially for the nonlinearities in the process, the valve plug is generally characterized; i.e., the flow versus stem travel characteristics follow a certain pattern. This also is an approximation since these characteristics deviate largely from any ideal curve. Thus, lack of accuracy exists because of non-ideal valve characteristics as well as because of non-ideal and non-constant process characteristics. It is quite obvious that under such conditions it is not realistic to look for excessive linearity and freedom from hysteresis in valve positioning or in any other control system component.

Nonlinearities, however, change the ratio between corrective action and deviation of the controlled variable. This is equivalent to changes in the sensitivity of the controller. Conceivably, nonlinearities may become so high that at certain valve positions, the system becomes unstable or corrective action becomes too slow.

Characteristics have been described in this chapter that refer to the output-input relations of components in control systems. The conditions investigated were those that exist *after* the system has undergone a change of input and has returned to the balanced steady-state condition. No attempt has been made to describe the behavior *while* the change is taking place. This will be discussed in the following chapters.

3. STEP FUNCTION RESPONSE OF A PROCESS

To determine the characteristic dynamic behavior of a process or control component, an input is artificially applied, and the output response that follows is observed. For the purposes of comparison, the input signal that is applied must follow a defined pattern. In Figure 3–1, three typical patterns of input signals are shown: The *step function,* the *ramp function,* and the *sinusoidal function.*

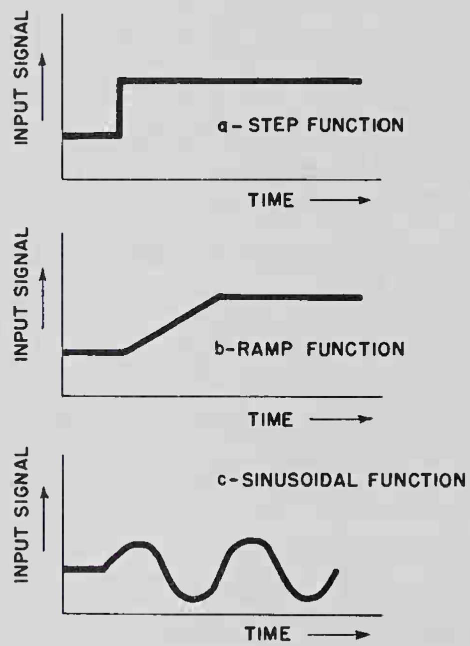


Figure 3–1. Patterns of input signals.

The ramp function, which corresponds to a gradually changing input signal, is hardly used today in control system analysis, though it may give interesting information regarding the response of an element. The analysis of a proportional plus reset plus rate controller in Chapter 4 does make use of it, however. The preferred approaches are step function and sinusoidal (i.e., frequency) methods which will be discussed in this and the next two chapters.

Methods and Results

The equipment necessary for determining the step function response of a process includes a means for recording the controlled variable and

a provision for manual loading of the final control element, together with an automatic-to-manual transfer switch. If the test is made under operating conditions, load changes must be eliminated for the duration of the test and a limited change of the set point during the test must be permissible.

An arrangement is illustrated in Figure 3–2, in which the controlled variable is pressure. The controller is disconnected and the final control

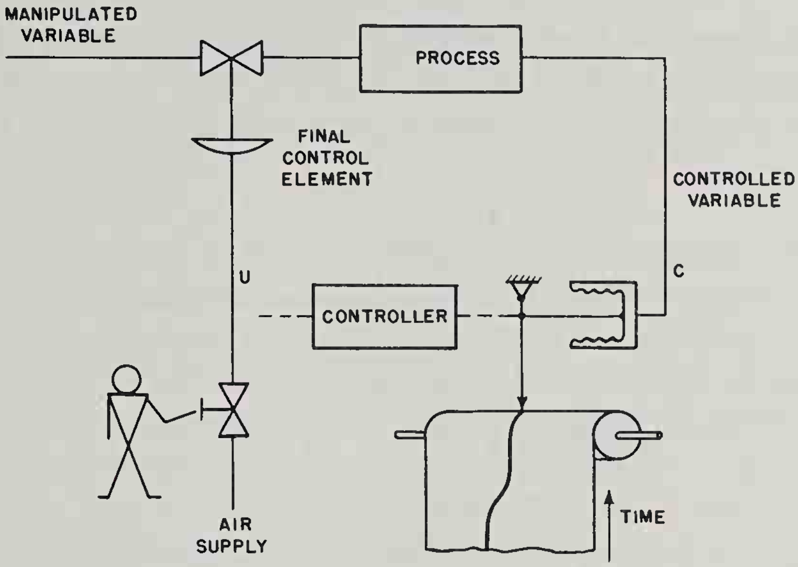


Figure 3–2. Arrangement of step-function response.

element is controlled manually. A sudden manual change of the air pressure applied to the final control element results in a change of the process pressure which is recorded and called the *step function response.*

The step change of the valve position should be as small as possible. The reason is that the interpretation of the step function response is based on the assumption of linear responses. Generally, this holds true only for sufficiently small step changes. Furthermore, the nonlinearity over the operating range may be so considerable, that the response for a valve which is, e.g., ¼ open is entirely different from the response for the same valve ¾ open. Hence, the step change should be repeated at various valve positions if the conditions of the installation so allow.

The step change varies the energy which enters the process through the manipulated variable. This energy change shows up in the process as variations in potential and kinetic energy, but also as a change of energy losses which is inherent in the majority of processes. The step function response may be considered a record of the method by which additions and losses of energy in a process take place.

It is possible to read from the step function response of a process its three basic characteristics:

1. The dead time.
2. The time constant as a result of resistances and capacitances.
3. The process gain as discussed further below.

With these data it can be determined what control action is best suited and for what values the various adjustments of the controller should be set.

Two basic forms of step function responses are illustrated in Figure 3–3. These are highly idealized patterns but they do contain the main components of practical responses. Figure 3–3a represents the step input,

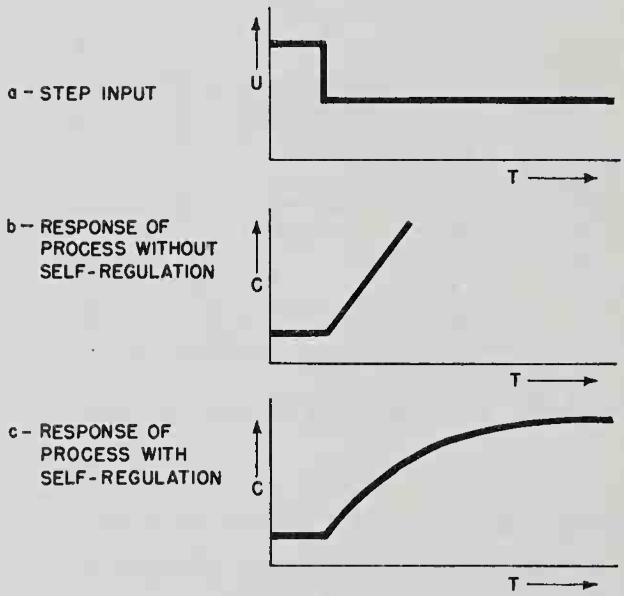


Figure 3–3. Patterns of process response.

i.e., the sudden manual change of the signal which under normal operating conditions is the controller output signal. Figures 3–3b and 3–3c are two basically different forms of responses to the step input as recorded from the controlled variable.

Process without Self-regulation

The response illustrated in Figure 3–3b is typical of a process which stores up potential energy at a constant rate without altering the energy losses that may occur as an inherent characteristic. Figure 3–4 is an example showing a tank with a liquid level control system. A pump maintains the output flow constant and a valve controls the inflow. Suppose the level is initially at equilibrium. A step change at the valve increases

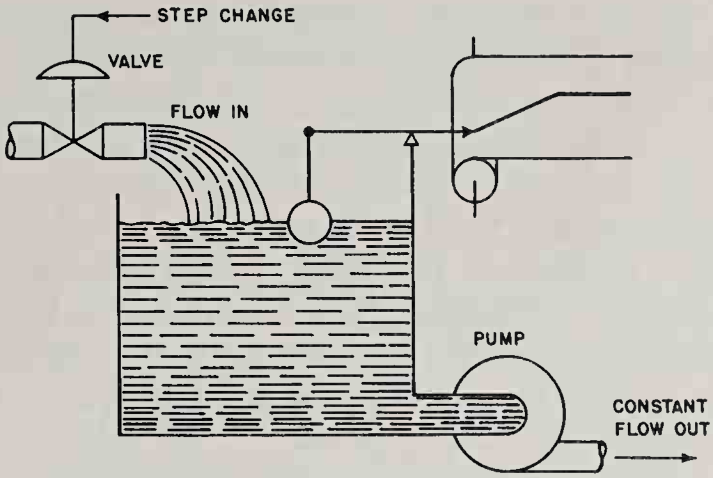


Figure 3–4. Process without self-regulation.

the input and the level rises. Because the output remains constant, the level continues to rise and a new balance is never reached. A process that responds in this form is said to have no self-regulation.

The rate of response is expressed by the ratio of the response to the time in which this takes place. Since C is the magnitude of the controlled variable, a change by a certain amount, e.g., from C_1 to C_2, is conveniently expressed by c. The time interval between T_1 and T_2 which is required to produce the increment c is expressed by t. The rate of response for a process without self-regulation is therefore expressed by

$$v = \frac{c}{t}$$

Process with Self-regulation

The response in Figure 3–3c differs from that seen in 3–3b in that it gradually approaches a new value of C. Such a response would correspond with the arrangement of Figure 3–5. Liquid enters the tank at a constant flow

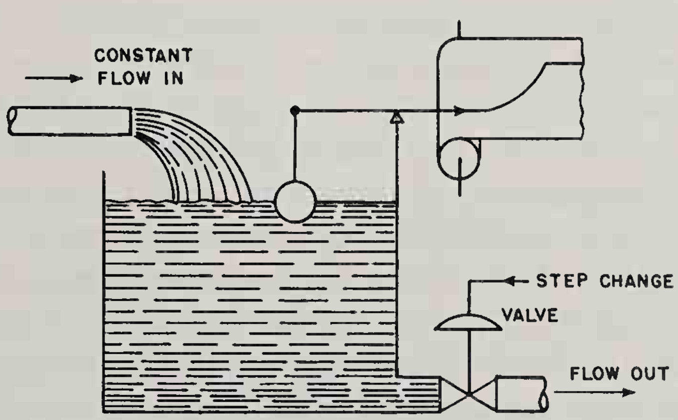


Figure 3–5. Process with self-regulation.

rate. The rate at which the liquid leaves the tank depends on the valve opening and the level of the liquid. Assume that Figure 3–3c represents the change in level seen in Figure 3–5 after a step change in valve position, introduced by manual means, closes the valve slightly. The inflow now exceeds the outflow and the level in the tank rises; but as the level rises, the hydraulic head increases also. The increase in head causes the outflow to increase. Eventually, at this new valve position, the inflow again balances the outflow. Thus a new balance has been reached even without the benefit of an automatic controller, although the offset may be considerable. The ability to reach a balance in this manner is known as self-regulation.

Another example is a temperature-controlled process. An increase of heat input is balanced by an increased heat loss to the surrounding atmosphere. Similar self-regulating conditions exist in the majority of industrial processes.

Gain, Response Rate, and Time Constant

The response curve of the self-regulated process in Figure 3–3c is determined by two conditions: the process gain and its response rate.

The gain is the total change in magnitude of the controlled variable per unit corrective action of the final control element. Suppose a step input applied to the final control element moves it by 0.1 in. and the pressure changes thereupon gradually until it settles at 20 psi above the initial value. The process gain in this case is 20/0.1 = 200 psi/in. valve motion. As an algebraic expression this reads

$$g = \frac{c}{u}$$

where g is the process gain, c is the increment of the process variable for a given step change of the final control element, and u is the magnitude of the step change.

The response rate is the maximum change of the controlled variable per unit time per unit corrective action of the final control element. Figure 3–6 illustrates the response of the same process for various magnitudes of step inputs. The maximum change per unit time is at the beginning of the response in all three cases. The speed is expressed by a line drawn tangent to the initial, i.e., maximum, slope of the curve. The steeper the tangent line, the faster the response. Suppose the step input moving the final control element by 0.1 in. results in a response with maximum change of 50 psi per minute of the controlled variable. The response rate in this case is 50/0.1 = 500 psi/min./in. valve motion.

Designating the process gain by g and the rate of response by v, the

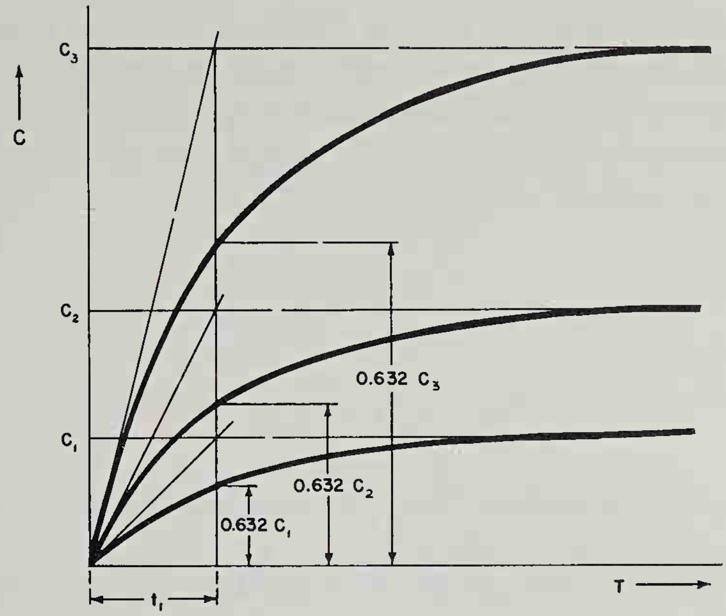


Figure 3–6. Step function responses with various magnitudes of step inputs.

two are combined in what is commonly called the time constant by writing

$$t_1 = \frac{g}{v} \tag{3-1}$$

Inserting the numbers of the examples for process gain and response rate, the result is a time constant of

$$t_1 = \frac{200}{500} = 0.4 \text{ min.}$$

Since it was shown that $g = \frac{c}{u}$, equation (3–1) may also be written:

$$t_1 = \frac{c}{uv}$$

Returning for a moment to a process without self-regulation, it is quite evident that its gain is infinite, hence its time constant is also infinite.

Various Definitions of Time Constant

The time constant is defined in many different ways. Its magnitude is always the same. Thus, one definition of the time constant is that it expresses the time which the controlled variable needs to pass through 63.2 per cent of its total change. This is illustrated in Figure 3–6. No matter how big is the step input for a given process, the controlled variable always reaches 63.2 per cent of its final value within the same amount of time which is given by the time constant t_1.

The tangents which are drawn in Figure 3–6 to the respective maximum speeds of response show that the ratio of the total change of controlled variable to t_1 is equal to the initial speed of response. It is only necessary to multiply this ratio by the magnitude of the step input u, to obtain the response rate of the process. Thus

$$u_3 \frac{c_3}{t_1} = v \tag{3–2}$$

and similarly

$$u_2 \frac{c_2}{t_1} = v \tag{3–3}$$

where u is used with suffixes 3 and 2 to indicate that different magnitudes of step change are involved.

However, the total change of the controlled variable divided by the magnitude of the step change is equal to the gain. Hence

$$\frac{c_3}{u_3} = \frac{c_2}{u_2} = g$$

Substituting this in either expression (3–2) or (3–3) gives

$$\frac{g}{t_1} = v \text{ or } t_1 = \frac{g}{v}$$

which is the same expression for the time constant as equation (3–1).

From Figure 3–6 it is also obvious that the time constant can be obtained from a line which is drawn tangent to the initial speed of response. The time at which this line reaches the final value of the controlled variable is the time constant.

Processes with Dead Time and Multiple Capacitances

Figure 3–7 compares a step function input signal with a process response in which a time interval t_2 elapses between input and the beginning of the response. This time interval is the *dead time* of the process.

Dead time is frequently used as an approximation in cases of multiple capacitance processes. It was mentioned in Chapter 1 that the time constant is the result of resistances and capacitances. There exists practically always a number of various resistances and capacitances in an industrial process. Occasionally one capacitance is considerably larger than the others, and for all practical purposes it is then justified to speak of a single-capacitance process. The response of single-capacitance processes corresponds to Figures 3–3 and 3–7. More frequently, however, two or three, and occasionally even more capacitances combine. The

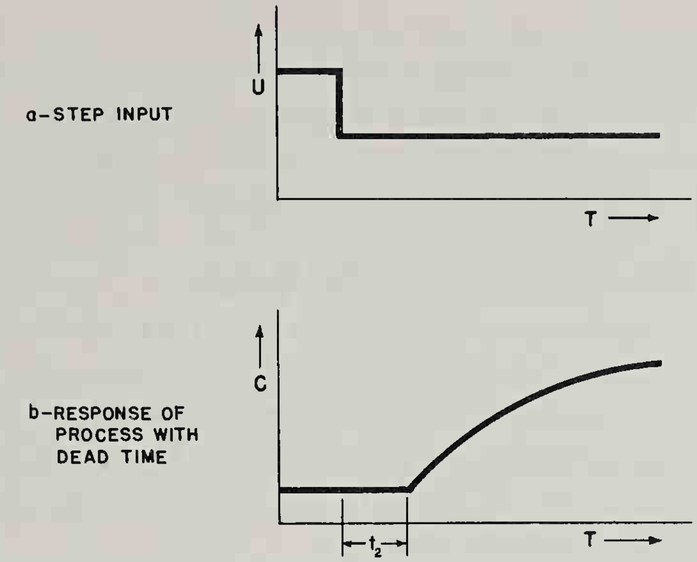


Figure 3–7. Step function response of process with dead time.

result is a composite S-shaped curve as shown in Figure 3–8. The sharp distinction between time constant t_1 and dead time t_2 is lost. The concept of these two time components, however, is so important for the evaluation of control systems that an approximation is used which proves to be sufficiently accurate. This approximation, together with some methods which are used in this text in slightly modified form to determine adjustments of proportional-position controllers, was first used by J. G. Ziegler and N. B. Nichols in their classic papers on "Optimum Settings for Automatic Controllers" and "Process Lags in Automatic Control Circuits," published in the ASME Transactions in 1942 and 1943, respectively.

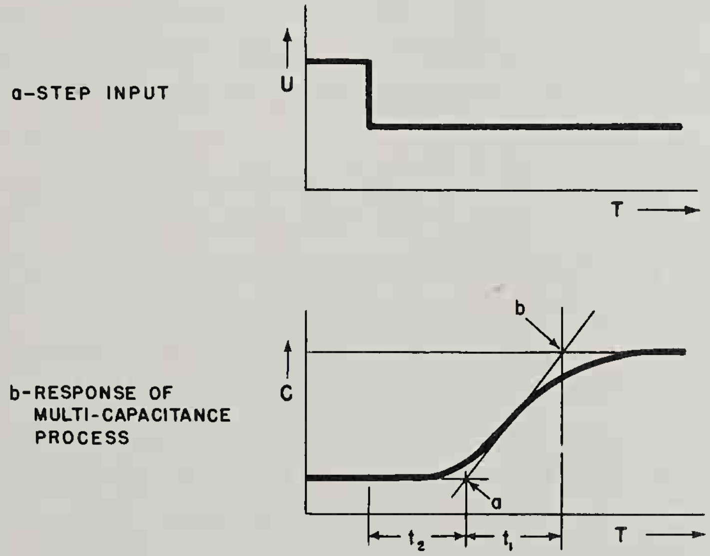


Figure 3–8. Response of multi-capacitance process.

It was mentioned that a line drawn tangent to the initial speed of response in Figure 3–6 intersects the final value of the process response after time t_1. The initial speed of response in this case was equivalent to the maximum speed of response. In the S-shaped curve of Figure 3–8, the maximum slope or speed of response is in the bend of the S, i.e., at the point of inflection. A line is drawn tangent to the response curve at the point of inflection. The points of intersection with the abscissas of the minimum and maximum values of the response curve are marked a and b, respectively. The time between the two points is considered the time constant t_1. The time between the step input and the first of the two points is the dead time t_2. These are the approximations which have proven to be very successful in their practical usage and will be used throughout this text.

Dead time is thus caused either by a transportation lag (see Figure 1–3) or by the additive effect of multiple capacitances.

4. STEP FUNCTION RESPONSE AND ADJUSTMENTS OF CONTROLLER

In order to utilize the data on process characteristics obtained by the methods described in the previous chapter, it is necessary to analyze the controller actions that may be combined with these processes. The most commonly used controller actions are:

1. Two-position action
2. Single-speed floating action
3. Proportional-speed floating action
4. Proportional-position action
5. Proportional plus reset action
6. Proportional plus reset plus rate action

In order to obtain a step function response from any controller, it is necessary to disconnect it from the final control element. In case of an electronic controller, an ammeter is connected in series with the specified load resistance across the controller output signal. For hydraulic controllers, a cylinder and piston of negligible inertia and friction replaces the final control element. The piston motion is picked up by a differential transformer, potentiometer, or similar device, to convert it into a signal for suitable recording. In pneumatic controllers, the controller output signal is applied to a pressure recorder. The medium measured by the primary element must be kept at constant conditions during the test. The step input is produced by a sudden small adjustment of the control-point-setting mechanism.

Figure 4–1 compares the arrangement for the step function test of the process with that of the controller. In Figure 4–1a the closed loop is shown connected for normal operating conditions. In Figure 4–1b, the loop is opened and between the step input and step function recorder are the final control element, process and measuring means. This means that this test does not investigate the process by itself but in combinations with the final control element and measuring means.

In Figure 4–1c, only the controller is inserted between step input and step function recorder. The two tests together comprise all the components of the loop. If the measuring means or final control element were combined with the controller instead of with the process, a pure approach

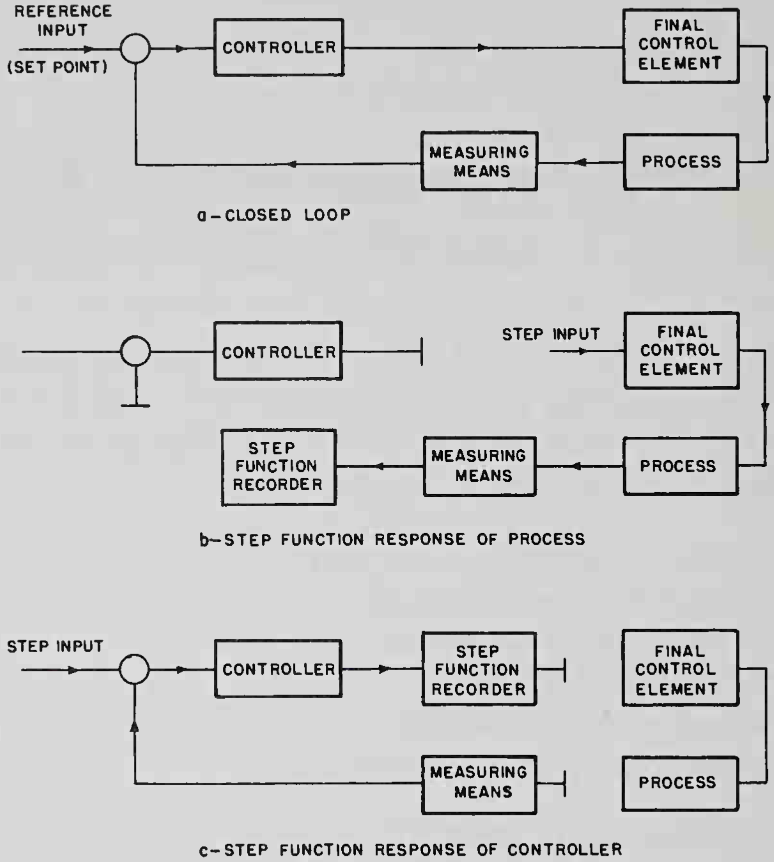


Figure 4–1. Step function tests.

typical of the controller action could not be obtained. This is the particular reason for the arrangements as shown.

In discussing the step function responses for various controller actions, two successive input steps as shown in Figure 4–2 will be considered. At time T_1 the first step is applied to the initial input. The second step follows at time T_2. This second step increases the input above the initial input by twice the magnitude of the first step change. An exception are two-position and single-speed floating controllers, for which the step inputs are inherently of the same absolute magnitude.

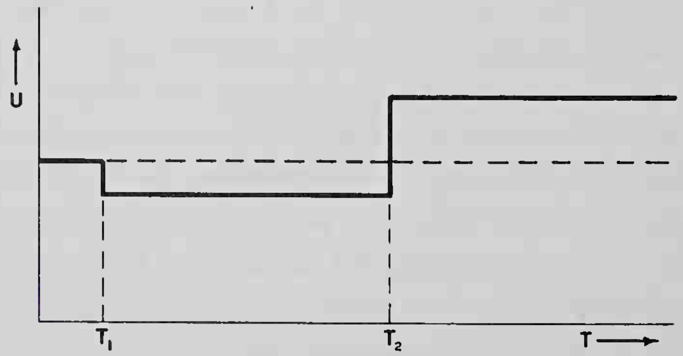


Figure 4–2. Step inputs for controller tests.

Two-position Action

Processes without Dead Time. Figure 4–3 illustrates a liquid-level control system. A solenoid valve controls the flow into the tank. The controller includes two metal rods *m* and *n*, which form part of an electric

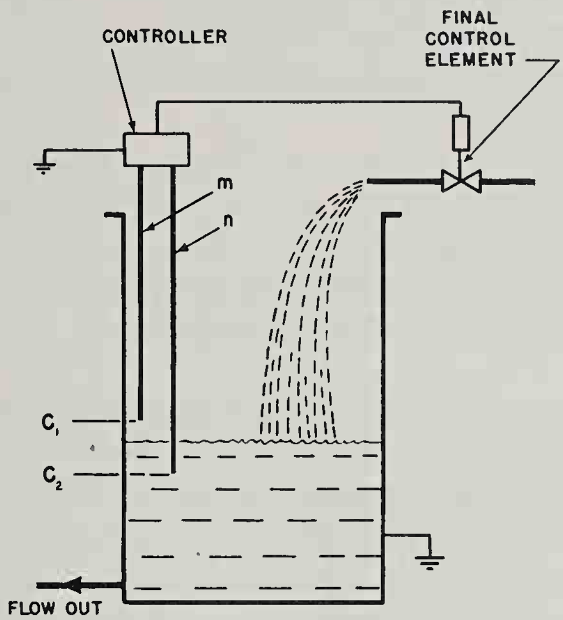


Figure 4–3. Liquid-level control system.

circuit that is closed through the ground connections as indicated. The liquid itself and the tank walls are electric conductors. In the condition shown, the level will rise until it reaches metal rod *m*. At this point an electric relay is energized and locked in, shutting down the solenoid valve. The valve remains closed until the level drops far enough to open the circuit through metal rod *n*. This unlocks the relay and *another* relay is energized and locked to maintain the valve open until the level rises again to *m*. There are only two input signals possible, either calling for flow or not calling for flow. The output, i.e., the tank level, is similarly limited to two alternative positions.

The process illustrated in Figure 4–3 has self-regulation since the flow out of the tank increases with the height of the level due to head pressure. With the same rate of liquid flowing in, the speed of response becomes slower as the level rises. The process response is hence comparable to Figure 3–3c.

Let the process be started up with an empty tank. The solenoid valve is fully open and the level rises along the process curve *a* illustrated in Figure 4–4. On reaching level C_1, the solenoid valve closes. The level now begins to drop along the inversed process response curve *b*. On

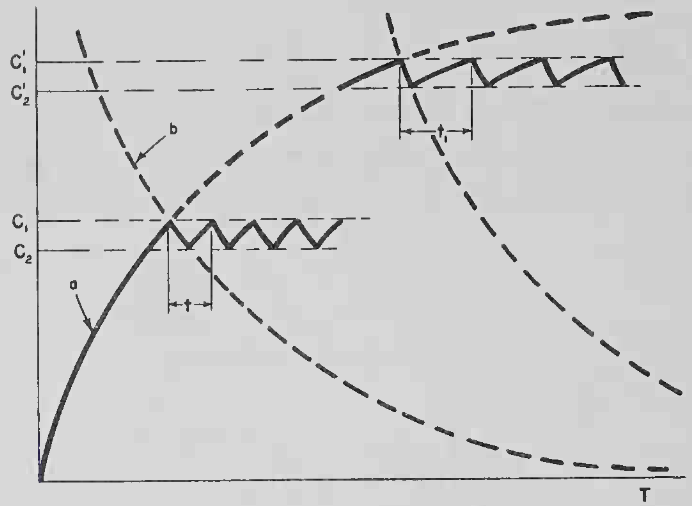


Figure 4–4. Process response for two-position control system without dead time.

reaching level C_2 the valve opens again. The difference between C_1 and C_2 corresponds to the difference in lengths between electrodes *m* and *n* in Figure 4–3 and is called the differential gap of the controller.

The smaller the differential gap of the controller, the shorter is the period *t* in Figure 4–4, and the greater is the number of times per minute of opening and closing the solenoid; that is, the frequency of its operation. How large this frequency can be is largely a question of life expectancy of the solenoid valve, contacts, and associate equipment. The frequency can be reduced for a given process by widening the differential gap. This means, however, that the level in the tank will oscillate between wider limits.

Another factor which determines the period *t* is the process response. A process with a large time constant, that is, a slow-responding process, will have a longer period *t;* i.e., the slower the process, the smaller can be the differential gap, and hence the closer are the limits within which a process can be controlled, for the same frequency of solenoid valve operation.

Since response curves *a* and *b* express the change of level for a fully open and closed valve, respectively, it can be deduced that the response would be slowed down by limiting the valve stroke to two extreme positions, e.g., ¼ open and ¾ open. This reduces the stroke to one half of full stroke and consequently reduces the flow variations caused by the controlling action. High-low or bypass adjustments are usually available with solenoid valves for this purpose. They allow the setting of any limits between which the valve will operate and thus make it possible to operate within a closer differential gap without changing the frequency of operation. The limitations of this method are given by the magnitude

of load change. The maximum consumption of liquid from the tank, for example, should not exceed the flow capacity at maximum opening, while the minimum consumption should not be less than the flow capacity at minimum valve opening.

The period t also depends on the set point of the controller. For example, controlling between levels C'_1 and C'_2, Figure 4–4, results in a longer period than between levels C_1 and C_2. In general, the longest period can be obtained by controlling toward either the upper or the lower level of the tank.

Processes with Dead Time. The conclusions for the two-position control system without dead time must be modified when dead time is present. As has been shown in the previous chapter, the concept of dead time also applies where the process consists of several capacitances connected by resistances. Such a process is illustrated in Figure 4–5. This

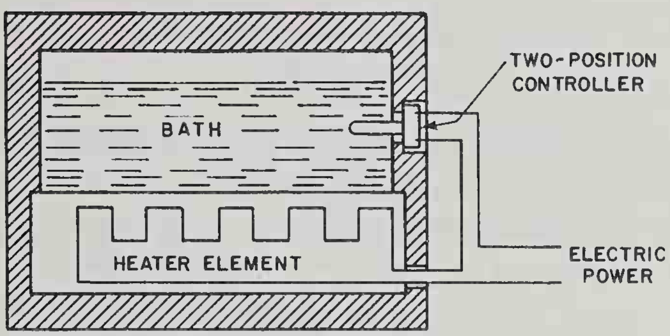


Figure 4–5. Two-position temperature controller.

is a temperature-controlled process. The temperature controller has electrical contacts like any thermostat that break when the temperature reaches a certain predetermined point. If the temperature drops by more than the differential gap, then the contacts connect again.

These contacts open and close the electric circuit that energizes the heater element, and thereby regulate the heating of the process. The electric current increases the temperature of the heater element which represents the first capacitance. The heat then flows through the resistance of the material in which the heater is imbedded and which constitutes another capacitance. The bath itself and the enclosure are further capacitances and resistances to be considered. All this results in a process with dead time.

The response is schematized in the graph in Figure 4–6. As shown, the process has been started up from a cold condition. If the contacts of the temperature controller were short-circuited, the temperature would rise along the dashed line up to a level which is determined by the self-regulation of the system. If the thermostat cuts off the heating

system at C_2, the temperature will still continue to rise somewhat due to the dead time which delays the effect.

If the time constant is sufficiently long, which means a more gradual slope of the response curve, Figure 4–6, then the overshoot is correspond-

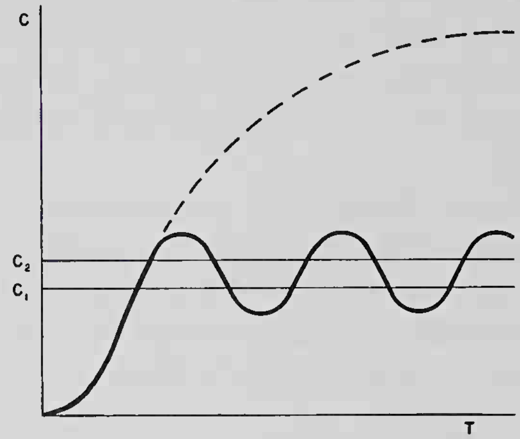


Figure 4–6. Process response for two-position control system with dead time.

ingly reduced and it is hence the *relative* length of dead time, or the ratio t_2/t_1, that is responsible for the overshoot.

The overshoot of a two-position controller is approximately

$$Y = \frac{gt_2}{2t_1}$$

where t_2 and t_1 are the dead time and time constant, respectively, and g is the process gain when the controller is switched from one position to the other. Thus a heat-treating furnace with on-off control may have a process gain of 2400°F, which means that at a room temperature of 100°F, the furnace temperature will increase to 2500°F when full heat is supplied. Self-regulation, i.e., increased heat losses, limit the possible temperature increase at this point and the process gain is hence 2400°F.

Assuming that this furnace has a dead time of 0.8 minutes and a time constant of 16 min., the overshoot is

$$Y = \frac{2400 \times 0.8}{32} = 60°\text{F}$$

If the furnace is to control at 1400°F, a recording thermometer will probably show that the temperature oscillates between 1340° and 1460°F. The heat-treated material itself, however, because of its relatively large capacitance, may maintain a comparatively constant temperature even under such wide fluctuations of the furnace temperature.

The above formula does not consider the contribution of the differen-

tial gap to the overshoot amplitude. Where $Y = 60°F$ it is quite obvious that the differential gap can be neglected, but where the magnitudes of Y are comparatively small and the differential gap can no longer be neglected, the amplitude will also depend on the differential gap.

There are several methods to reduce the overshoot amplitude. Where the differential gap contributes essentially to the overshoot, the most obvious step is to reduce the differential gap. This has, however, mechanical limitations due to backlash and lost motion. In a bimetallic thermostat, for example, it is hardly possible to reduce the differential gap below 1.5° F.

Another method is the accelerator, a heater element frequently built into thermostats, which is energized whenever the latter calls for heat. The primary element, usually consisting of a bimetallic spring, responds to the additional heat of the accelerator and therefore cuts off the heat source before the upper limit of the differential gap is actually reached. This is equivalent to reducing the differential gap and hence limits overshoot.

A third method, which is frequently used in heat-treating furnaces, consists in operating at a slower heat-producing rate, hence at a longer time constant, when approaching the set point. The higher rate is used for starting up to avoid delays and the lower rate reduces overshoot. A multi-position controller switch automatically changes from one rate to another at a pre-set temperature.

Load Changes. Figure 4–7 shows the effect of load changes by means of three curves, namely, the load curve, the curve that shows the

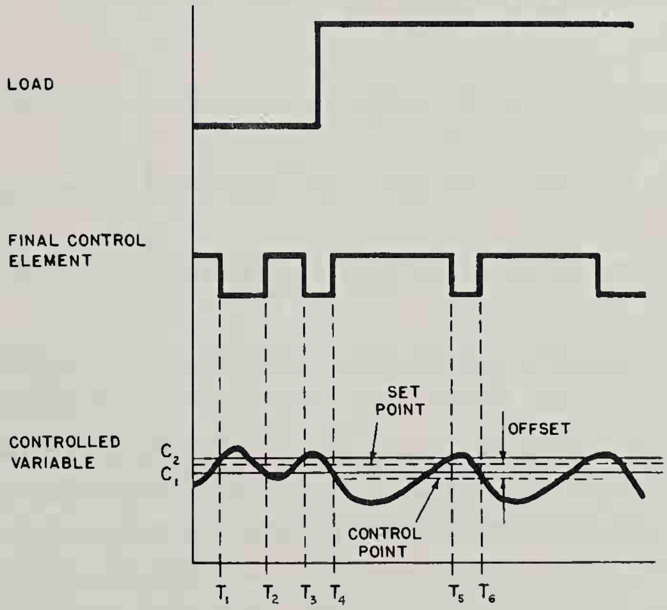


Figure 4–7. The effect of load changes on two-position control systems with dead time.

position of the final control element, and the curve with the fluctuations of the controlled variable. The controlled variable is assumed to be temperature and the process response is considered to include dead time. The final control element, for example, a valve in a steam heating system, is shut down when the controlled variable reaches the temperature C_2 at T_1; however, due to dead time there is some overshoot. When the temperature begins to drop, it has to reach C_1 before the valve opens and again there is some overshoot. The difference between C_2 and C_1 is the differential gap.

At T_3 the valve is closed again, but shortly after, a load change occurs that calls for considerably more heat input. Since the valve is closed, the temperature drops at a faster rate than before. This is equivalent to a shorter time constant, and results in a larger t_2/t_1 ratio, hence larger overshoot as expressed in the graph. The temperature finally builds up again, at a slow rate since the load continues to be large and the consequence is that the overshoot between T_5 and T_6 is small. Nevertheless, the amplitude of oscillations has increased considerably.

Another effect is that there is a shift of the center around which the controlled variable oscillates. This separates the control point from the set point. The center of the differential gap is considered the set point. Initially this coincides with the center of the oscillations. After the load change, as shown in Figure 4–7, the center of the oscillations has dropped as indicated by the control point. Originally, set point and control point are identical. After the load change, they are separated by the offset.

Conclusions. The following conclusions can be made about two-position action:

1. The controlled variable oscillates within a certain amplitude above and below the set point.
2. For a given process and load, the amplitude depends on the differential gap.
3. The minimum differential gap is generally limited by mechanical conditions.
4. In a process *without* dead time, the minimum differential gap is also limited by the maximum permissible frequency of operation of the final control element. The frequency increases inversely to the time constant of the process. Hence processes with short time constants are not suited for two-position action.
5. In a process *with* dead time, the overshoot is the limiting factor. For a given differential gap and load, the overshoot depends on the ratio of dead time to time constant (t_2/t_1) and on the gain g.

Hence processes with large t_2/t_1 ratio and large g are not suited for two-position control.

6. A number of remedies are available to improve the response of a two-position controller, such as high-low or bypass adjustments in solenoid valves, accelerators and multi-position controllers.
7. The difficulties caused by overshooting increase with load changes.

In spite of all limitations, there are a great number of processes in the industry that allow two-position action which results in the most simple and generally the most economical controller.

Single-speed Floating Action

Figure 4–8 illustrates a single-speed floating controller. The controlled variable is liquid level head which is applied as pressure to the bellows.

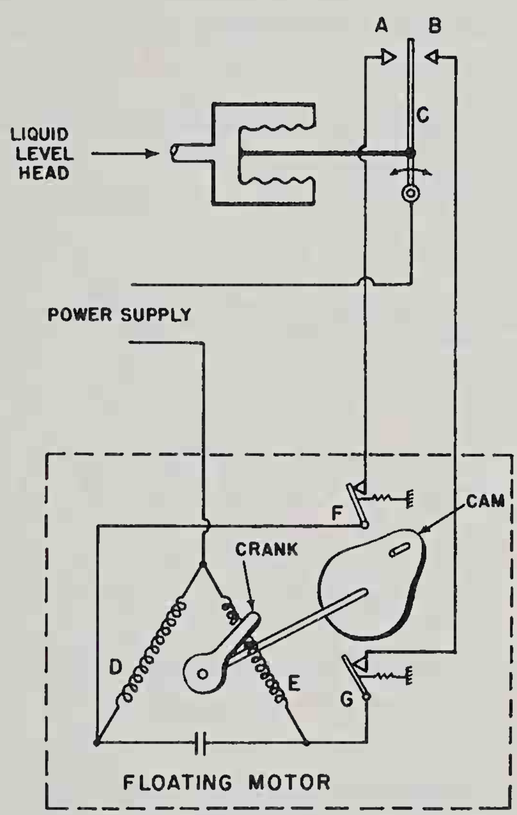


Figure 4–8. Single-speed floating controller.

Increase of level, i.e., pressure, and hence expansion of the bellows, will result in contact between B and C. Inversely, decrease of level produces contact between A and C. A certain amount of level change may occur around the midposition of C without making contact on either side. This range of pressure change is the neutral zone.

The motor is a shaded-pole reversible single-phase motor. If winding

D is energized, the crank and cam rotate counterclockwise; if *E* is energized, clockwise. The travel is limited, e.g., to 180 degrees, by means of the cam which is mounted on the motor shaft and actuates limit switches *F* and *G*. For example, if contact *C* closes on *A*, winding *D* energizes, and the motor rotates counterclockwise until it opens contact *F*, deenergizing winding *D*. No further action occurs until contact *C* closes on *B* and starts the motor in clockwise rotation. The crank rotating on the motor shaft would be linked to a final control element.

Processes without Self-regulation. Suppose the controller positions a valve in the water line to a boiler drum. If the boiler drum is of equal cross-sectional area throughout its height and the valve is suddenly opened beyond the requirements of the process, the level rises continuously. This characteristic defines a process without self-regulation.

Let the system be initially in balanced condition. This is shown in Figure 4–9 for the case of a controller without neutral zone. Before

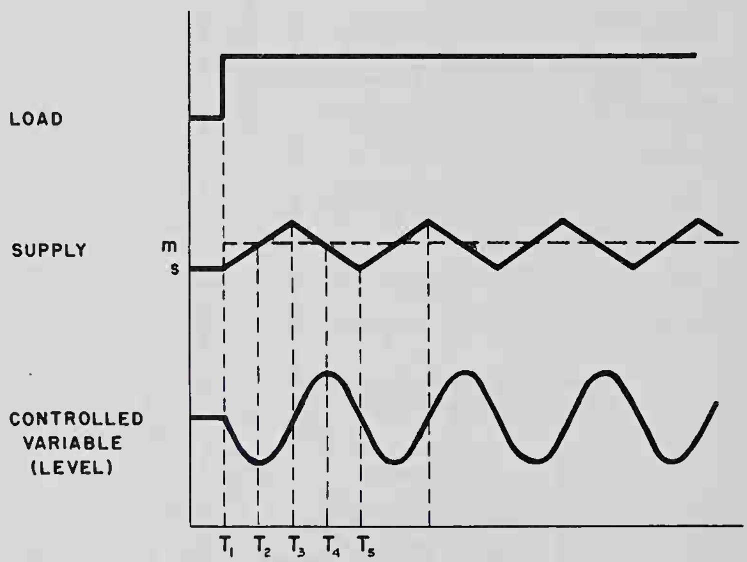


Figure 4–9. Process without self-regulation with floating controller (no neutral zone).

time T_1, a given load corresponds with a certain valve opening which assures enough water supply to maintain the controlled variable, i.e., the level at the set point. At time T_1 a load change enters, the steam demand increases and the level drops. From Figure 4–8 it can be seen that as a result of this change, contact *C* moves on *A* and energizes the motor, which opens the valve to increase the supply as shown in Figure 4–9. While the full line *s* corresponds with the supply flow into the tank, the dashed line *m* represents the rate of liquid removed in form of steam. The valve opens until the supply *s* equals the demand *m*, which occurs at time T_2.

However, at this point it is not enough that supply equal demand. As long as the deviation from the set point persists, added supply is required to make up for the deficit in liquid level. The valve will therefore continue to open until T_3, at which time the controlled variable recovers the set point. At this point, the valve is too wide open, because the excess of supply to return the level to the set point is no longer needed. The valve begins to close. Until T_4, however, the level continues to rise, since only then has the valve obtained the opening m which would deliver the required quantity; however at this time the level is too high. Hence the valve goes on reducing the flow until the required level is reached at T_5. At time T_5, however, the situation is the same as it was at time T_1, namely the supply is not sufficient for what the load demands. The cycle starts again.

As soon as dead time is added to these conditions, the response gets worse. Dead time may exist in the process as well as in the controller. In the following example the controller is assumed to be the cause of the dead time, e.g., because the motor of Figure 4–8 coasts slightly before it reverses. The effect of such coasting is illustrated in Figure 4–10. At time T_1 the load increases and the controlled variable begins to drop.

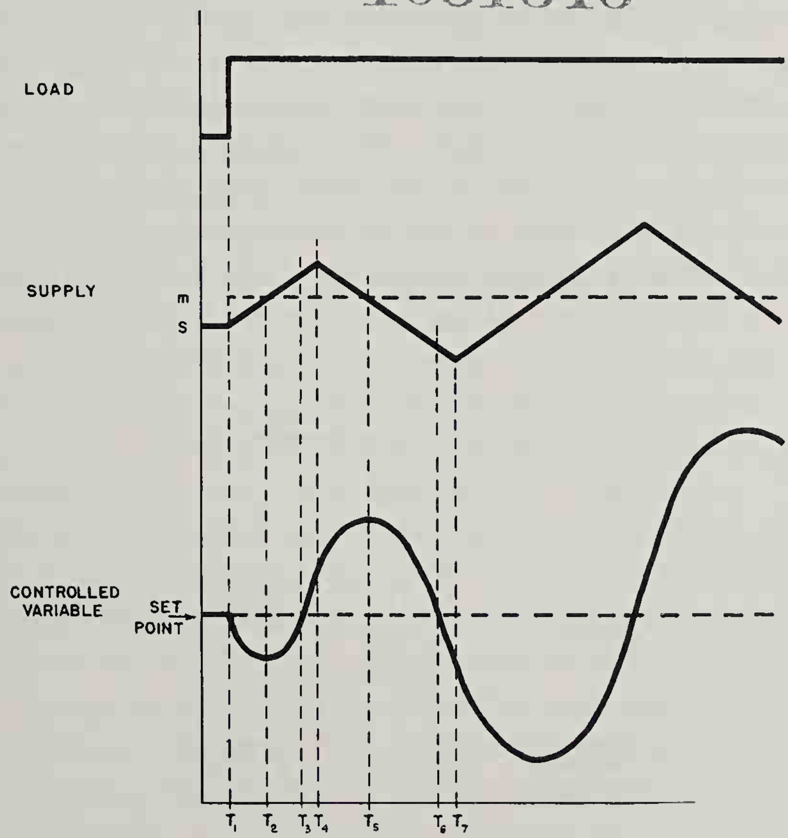


Figure 4–10. Process without self-regulation with floating controller (actuating motor coasting).

Action of the controller gradually increases the supply s reaching the m-line at T_2, from which time the controlled variable begins to return to the set point which it reaches at T_3. At this point, the motor should reverse but due to coasting it continues its movement to T_4. The m-line is now crossed at T_5, and an increase of amplitude in the response of the controlled variable is already visible. From T_6 to T_7 the motor coasts again before it reverses and this results in a further amplitude increment. It is obvious that the controller will rapidly reach its maximum amplitude, limited by mechanical conditions, and will continue to oscillate within these amplitude limits. Such control is undesirable, and single speed floating action can therefore not be used with processes without self-regulation.

Processes with Self-regulation. The above described situation changes completely when the process contains self-regulation. A typical example is an air-conditioning system.

Suppose the motor actuator of Figure 4–8 positions the damper in a duct admitting hot air to a recirculating drier. The drier temperature is the controlled variable and the bellows in Figure 4–8 may be considered gas-filled and sealed. On a drop in temperature the gas contracts, the bellows moves with it, and contact C moves toward A.

If the controller were to be disconnected, and the damper were manually changed, e.g., toward closure, then the temperature would drop. This drop would, however, not continue indefinitely as in processes without self-regulation, but would level out at a new temperature. This is because with the decrease of the drier temperature, the heat losses through the ducts and walls of the drier also become less.

Figure 4–11 illustrates the action of such a system. Dead time is neglected. At time T_1, a sudden increase in heat demand offsets the

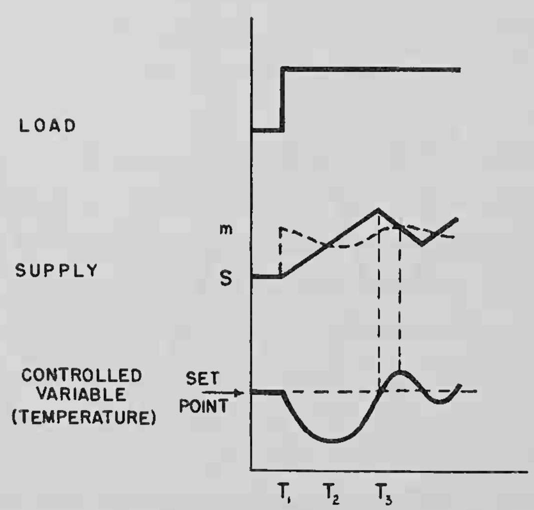


Figure 4–11. Process with self-regulation with floating controller.

balance. The temperature starts to drop and this leads to a gradual opening of the damper and an increase of warm air supply.

As in previous examples, the *s*-line expresses the position of the final control element, and simultaneously the flow into the controlled system, while the *m*-line interprets the flow leaving the system, which in the present case is the heat loss from the drier system. This flow leaving the system is no longer a constant as in Figures 4–9 and 4–10, but becomes a function of the controlled variable.

When the temperature drops, the heat loss diminishes, which is expressed in the *m*-line. Thus, the *m*-line reflects the oscillating response of the controlled variable. The crossover of the *s*-line and *m*-line indicates the point where the supply is equal to the heat loss. This happens at T_2, and from this moment on the temperature begins to rise. At T_3 the set point is reached and the damper begins to close. The fact that the *m*-line is no longer straight but is deflected in a direction that shortens the time until m equals s, having thus a damping effect on the dynamic action, leads to rapidly subsiding cycles. The control system is stable, which proves that the process with self-regulation lends itself to control by floating action.

Figure 4–11 also illustrates the fact that the greater the speed of the damper motor is, the sooner will the *m*-line be reached, and the smaller will be the amplitude of the temperature oscillations. This is true for an ideal system, but dead time and time constant must again be taken into account for practical applications. The magnitude of dead time produces overshoots for reasons quite similar to those illustrated in Figure 4–10. The difference is that in the damped response of Figure 4–11, the resulting oscillations are not necessarily of increasing amplitude, but may rapidly subside.

An offset as previously described for the two-position controller cannot exist with a floating controller. As long as a deviation of the controlled variable exists which is larger than the neutral zone, the floating controller will produce corrective action.

The floating action of the controller provides a gradual corrective action. The faster the process response is, i.e., the shorter the time constant, as compared to the floating speed of the controller, the less necessary will it be to move the final control element in order to obtain a measurable change in the controlled variable, and hence the less will be the amplitude of the overshoot. For the same reasons, the time constant of the measuring element should be small as compared with the floating speed of the controller.

This implies rather serious limitations on the floating speed and hence on the speed of correction when load changes occur. As long as these

load changes proceed slowly, the slow corrective action is of minor importance; however, if these changes are fast, then the time that elapses before the controlled variable is returned to the set point may become objectionably long.

Conclusions. The following conclusions might be made about single-speed floating action:

1. Single-speed floating action can be used only with processes with self-regulation.
2. The smaller the dead time and the time constants, the better the control.
3. The floating speed, and hence the speed of corrective action, has to be slow, relative to the time constants of the process and the measuring means, to keep the amplitude of the oscillations within acceptable limits.
4. The relative slow corrective action permissible with this type of control action becomes more obvious and hence objectionable in processes with fast load changes.

Proportional-speed Floating Action

With this action the speed is proportional to the deviation of the controlled variable from the set point. The action of such a controller can be faster by far than that of a single-speed floating controller, because in reducing the deviation of the controlled variable and bringing it back to zero, it automatically slows down the floating speed.

It is used most frequently with hydraulic controllers and, hence, in those applications where not only the advantages of proportional-speed floating action but also the high power output of hydraulic actuators is desirable, such as in large dampers, heavy butterfly valves, etc.

Figure 4–12 illustrates the action of a proportional-speed floating controller. Process pressure is the controlled variable. It is applied to a diaphragm. The deflection of the diaphragm depends on the magnitude of the process pressure and the spring constant of the spring which opposes the force exerted by the process pressure. The spool of a four-way valve is connected to the diaphragm through a lever system. The result is that the spool position is proportional to the process pressure. The spool controls the oil flow to the actuating cylinder. The position shown persists as long as the controlled variable is at the set point. If the process pressure rises, the spool moves downward. This connects the lower part of the cylinder with the oil supply and the upper part with the drain. As a consequence the actuating piston moves upward. Inversely, a decrease in process pressure results in a downward motion of the actuating piston.

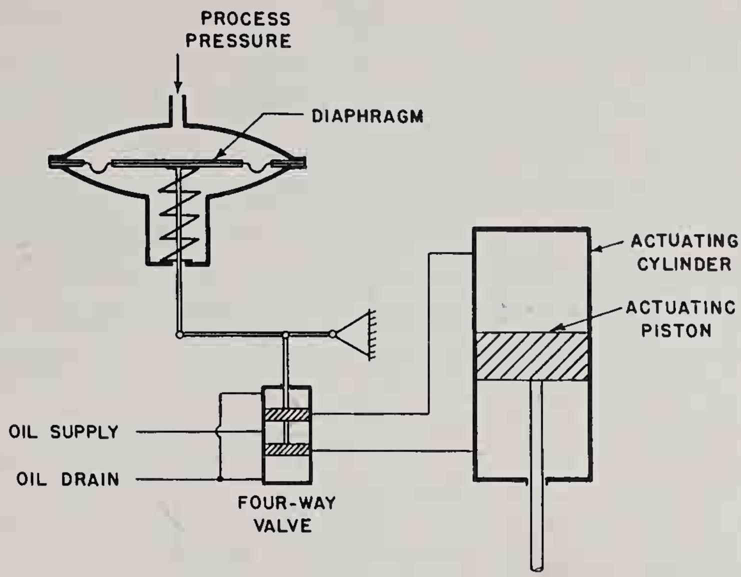


Figure 4–12. Proportional-speed floating action.

The piston is linked to a final control element which it positions accordingly.

The speed of the actuating piston depends on the rate of flow of oil passing through the four-way valve. The rate of flow is a function of the valve opening. Since the opening of the four-way valve increases with the deviation of the controlled variable from the set point, it follows that the actuating piston moves faster, the larger the deviation.

Thus the two characteristics of proportional-speed floating action are illustrated. The motion of the final control element continues as long as the deviation of the controlled variable from the set point persists. The speed at which the final control element moves is proportional to the deviation of the controlled variable from the set point. There are, of course, limitations both to duration and speed of the motion of the final control element, since the motion will discontinue when the end of the stroke is reached, and the speed cannot be faster than the oil supply allows.

Calling x the displacement of the final control element and t the time required for this displacement, then the ratio x/t expresses the floating speed. It is proportional to the deviation e of the controlled variable from the set point. Hence

$$\frac{x}{t} = fe$$

where f is a constant which is called the floating rate of the controller.

Figure 4–13 compares the response between single-speed and proportional-speed floating action, illustrating for the latter the gradual de-

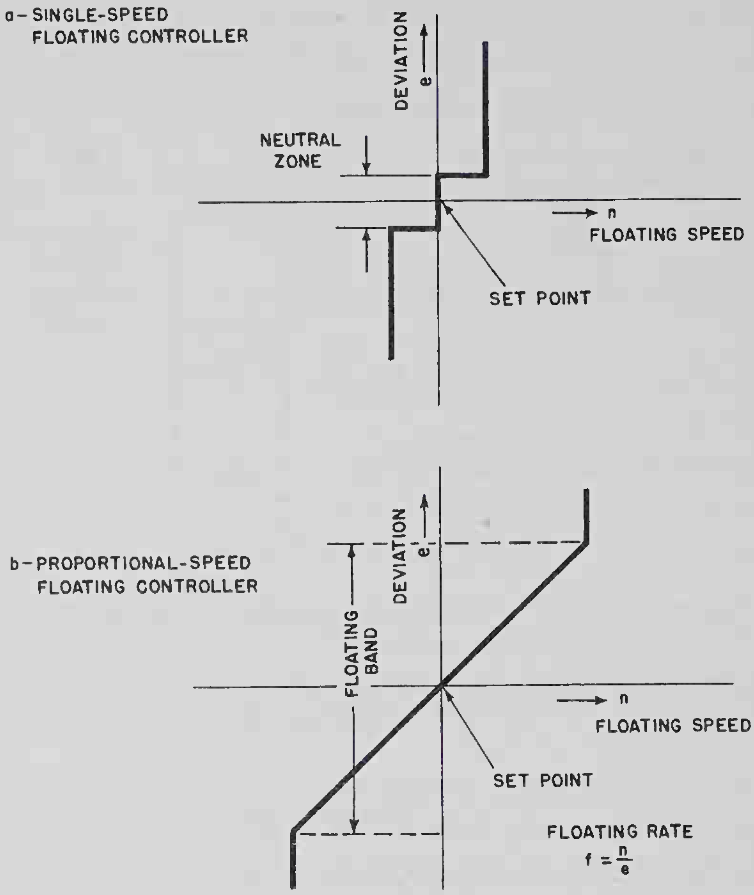


Figure 4–13. Floating speed vs. deviation of single-speed and proportional-speed floating controllers.

crease of floating speed as the controlled variable approaches the set point, as compared with the abrupt change of the single-speed type.

The speed x/t of the final control element is expressed by n, and hence

$$n = fe$$

The floating speed n is limited by the maximum speed that can be obtained from the controller. From previous considerations, it is obvious that the floating rate should be as high as possible without sustained oscillations. Adjustments on the controller allow attainment of the optimum floating rate.

Adjustment of Floating Rate. A slow floating rate is identical with low sensitivity of the controller, and vice versa. Figure 4–14 shows several responses of a controlled variable, assuming that initially at $T = 0$, the set point is changed from C_1 to C_2. The slowest floating rate, i.e., lowest sensitivity, would result in an overdamped response as shown. The floating rate can be increased to a maximum, a point at which still

no overshooting of C_2 occurs. This is called *critical damping* and expressed by the corresponding curve in Figure 4–14. Any further increase of floating rate results in oscillations. The ¼-decay ratio means that each successive overshoot is ¼ the magnitude of the previous one. Similarly, the ½-decay ratio reduces each overshoot to ½ of the previous one. In the interest of fast corrective action and limited overshoot, the ¼-decay ratio is generally the preferred one. In a single-capacitance process, for example, the ¼-decay ratio allows operation more than 20 times faster than under critically damped conditions.

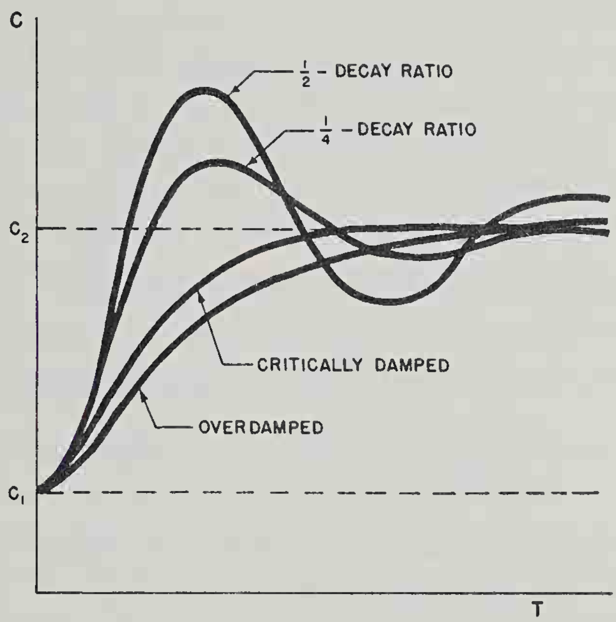


Figure 4–14. Various degrees of damping.

The floating rate for ¼-decay ratio for a process without dead time is given by

$$f = \frac{5}{g t_1}$$

where f = floating rate of final control element in in./min./unit change of controlled variable

g = process gain in units of controlled variable/in. displacement of final control element

t_1 = time constant in min.

If dead time t_2 is present, or if a multi-capacitance process is broken down into a single time constant plus a dead time, allowance has to be

made correspondingly. A practical method with sufficient accuracy for most applications is based on the equation

$$f = \frac{k}{g(t_1 + t_2)}$$

where k is a number chosen from the following table:

t_1/t_2	k
greater than 25	5
12–25	4
6–12	3
2–6	2
1–2	1.5
smaller than 1	1

For example, a temperature-controlled process has a gain of 5°F/in. of valve motion, dead time of 1 min. and time constant of 3 min. Substituting values in the above equation gives a floating rate of

$$f = \frac{2}{5(3 + 1)} = 0.1 \text{ in./min./°F}$$

A method for determining the floating rate in an actually installed system consists of the following steps:

1. Increase of the floating speed to the critical setting where the control system just begins to oscillate without changing amplitude.
2. Measurement of the time needed per unit length of stroke of the final control element under these conditions.
3. Estimating the relation of time constant to dead time (t_1/t_2) and decreasing the floating rate so that the time measured in the second step is increased by the factor listed in the following table:

t_1/t_2	Increase of Stroking Time
25–100	10 times
10–25	5 times
4–10	3 times
smaller than 4	2 times

The table gives no values beyond $t_1/t_2 = 100$, since beyond this value it is generally not possible to obtain continuous oscillations.

Limitations for critical setting. Rules for critical setting as given for the different mechanisms in this chapter require certain qualifications. The response of a process to a load change depends on the nature of the load change, and where load changes of different origins are expected, all of them should be investigated. On a feedwater heater, for example, the temperature may be influenced by the amount of throughput as well as by the steam pressure. Hence the critical setting should be repeated by keeping the set point constant and changing the steam pressure slightly. The setting should be determined by whichever method requires less sensitivity.

Furthermore, nonlinearities may exist, e.g., between the change in position of the final control element and the change in magnitude of the controlled variable. Valve characteristics are used to make this relation at least approximately linear. If the nonlinearities are too high, the control system may be stable at one valve position but unstable at another. The answer is to change the valve characteristic or to reduce the sensitivity of the controller sufficiently to accommodate a wide enough range of valve openings.

Proportional-position Action

In the proportional-speed floating controller described above, the position of the final control element is independent of the deviation of the controlled variable from its set point. As long as the deviation persists, the final control element continues to move—as far as its mechanical limitations allow. In the proportional-position controller, on the other hand, the response becomes a function of the deviation of the controlled variable.

Figure 4–15 shows the responses of different controllers. A similarity exists between the two-position and the proportional-position controller insofar as the response is an immediate one in either case, and does not correct gradually as in the floating controller.

The corrective action caused by the proportional-position controller is modified by the final control element. A spring-opposed diaphragm valve does not respond immediately, but a certain time is required to move it from one position to another. Figure 4–16 compares the proportional-position action of the controller with that of the controller and the final control element combined. This shows that the final control element becomes the limiting factor in the fast corrective action of a proportional-position controller.

Proportional-position action is used equally in electric, electronic, hydraulic, and pneumatic controllers. Figure 4–17 illustrates a pneumatic application. The signal from the controlled variable is applied as pres-

sure to the signal bellows. An increase of pressure will tend to rotate the flapper clockwise. This brings the flapper closer to the nozzle, decreasing the opening through which air flows out.

Supply air passes through a fixed restriction and then through the nozzle to atmosphere. Due to the movable flapper the nozzle tip becomes a variable restriction. The air flow passes through two successive

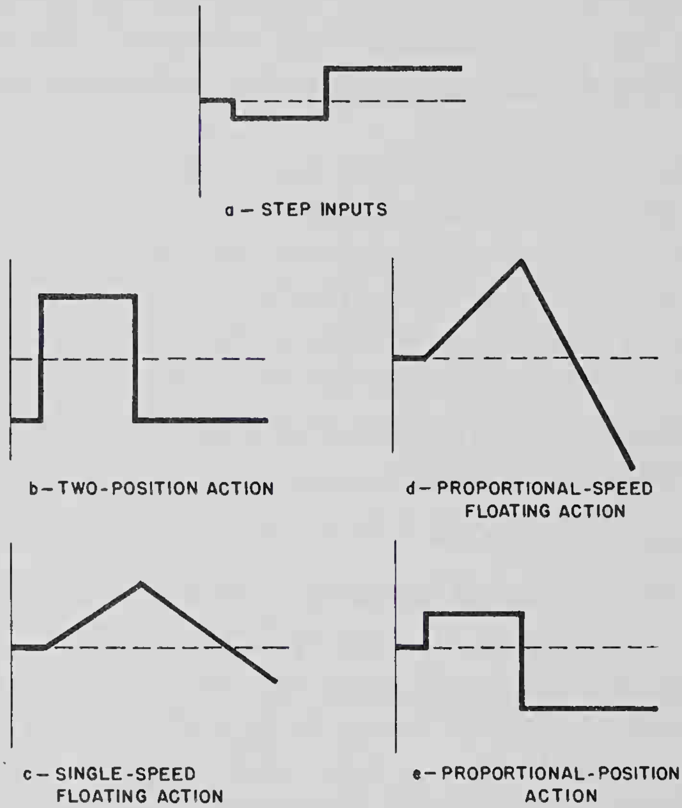


Figure 4–15. Step function responses of various controllers.

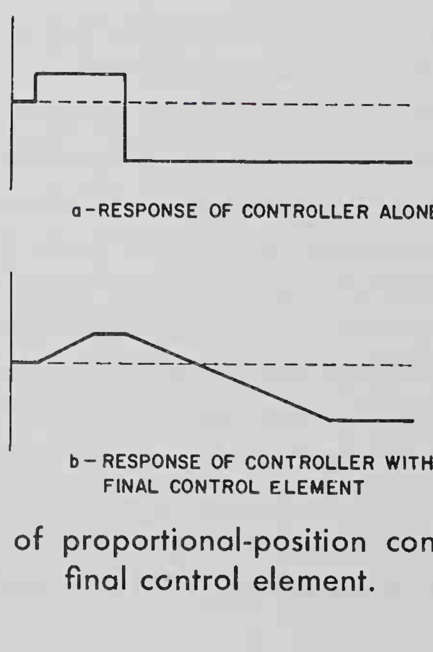


Figure 4–16. Response of proportional-position controller without and with final control element.

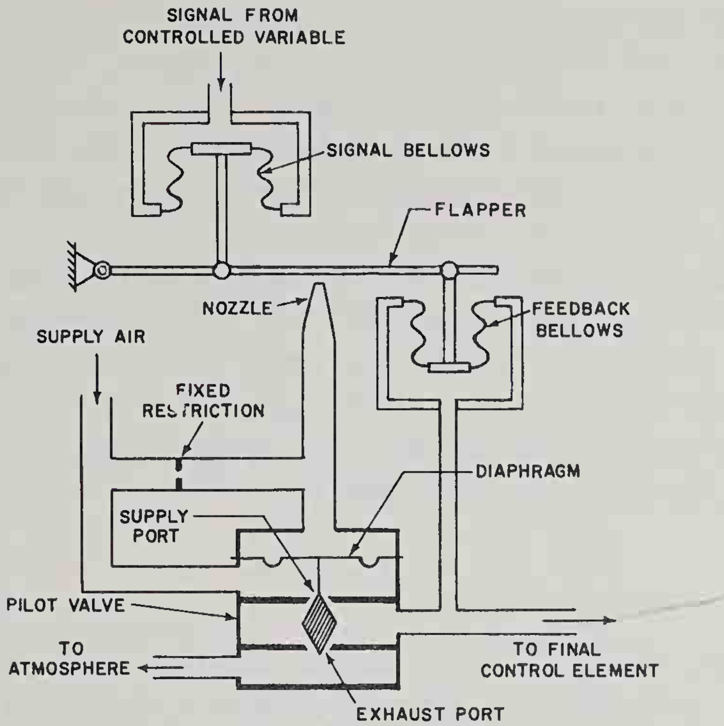


Figure 4–17. Proportional-position controller (pneumatic application).

pressure drops, one in the fixed restriction, the other in the variable restriction. The pressure between the two restrictions is the nozzle back pressure, and it will assume a value determined by the flapper position. When the flapper approaches the nozzle, the nozzle back pressure increases. When it moves away, it decreases.

Since relation between flapper position and nozzle back pressure is linear only within relatively small pressure changes and since, furthermore, the flow capacity that can be handled by such an arrangement is rather limited, it is customary to use a pilot valve which increases the effect of the flapper-and-nozzle arrangement and which regulates the air flow to and from the final control element.

The nozzle back pressure is applied to a diaphragm in the pilot valve. Connected to the diaphragm is the plug of a three-way valve, which connects the final control element through the supply port to the supply air and through the exhaust port to atmosphere. When the supply port tends to close, the exhaust port opens, and vice versa. Again there are restrictions connected in series, the supply port and the exhaust port. Both restrictions are variable, and the pressure that exists between them is applied to the final control element.

Suppose the signal from the controlled variable rotates the flapper clockwise, the nozzle back pressure increases, the exhaust port closes, the supply port opens, and hence the pressure between the two ports in-

creases. This increased pressure is transmitted not only to the final control element, but also to the feedback bellows, which now exerts a counter-clockwise force on the flapper. This is the picture of a force balance between signal bellows and feedback bellows. Their balanced position determines the deflection of the flapper and hence the signal transmitted to the final control element.

If the final control element is a spring-opposed diaphragm valve, as it usually is, this pressure is converted into a force and balanced with the force of a spring. The result is a finite position of the final control element which is proportional to the signal from the controlled variable.

The step function response of a proportional-position controller as illustrated in Figure 4–15e is expressed by

$$x = \frac{100}{P}e$$

where x is the change of the controller output signal, e is the magnitude of the step input, and P is the proportional band in per cent of instrument range. The meaning of the proportional band is discussed in detail further below.

Offset. The most important limitation of a proportional-position controller is its inability to return the controlled variable to its set point after a load change has caused a deviation. This is illustrated in Figure 4–18. The regulation of the supply, indicated by the center curve, corresponds to the output of the controller and final control element combined. The change in controlled variable expressed by the lower curve represents the input to the controller. Since in the proportional-position controller the output is inversely proportional to the input, their curves must correspond with each other, one being the mirror image of the other.

After the load change, the controlled variable begins to decrease and the supply rises. The purpose of the increase in supply is to bring the controlled variable back to its set point. However, this is impossible for the proportional-position controller, because its basis is that controlled variable and supply maintain their proportion, and the mechanism does not permit maintenance of one at the set point while the other operates at a higher level. The proportional-position controller can only reduce the deviation of the controlled variable from the set point, but cannot eliminate it. The remaining deviation is called the offset, as shown in Figure 4–18.

The extent to which the proportional-position controller reduces the deviation that would occur without it, is expressed by the deviation reduction factor Q. For example, a feedwater heater is to raise the feedwater temperature to 180°F. For a given steam admission

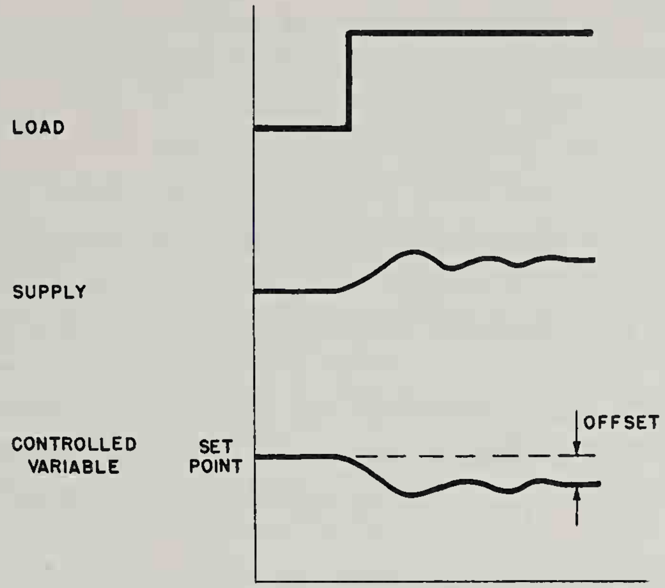


Figure 4–18. Proportional-position action.

this temperature can be maintained as long as a certain average feedwater demand from the boiler exists. However, at peak periods, the temperature would drop to 160°F unless an automatic controller is used. Let the controlled variable be the outlet temperature of the feedwater and the control agent be the steam. Suppose the control system is to prevent the feedwater temperature from dropping below 178°F. This means that without the controller, the feedwater temperature could decrease by a maximum of 20°F, with the controller the maximum decrease is reduced to 2°F. The allowable deviation reduction factor is then

$$Q = \frac{180 - 160}{180 - 178} = \frac{20}{2} = 10$$

Obviously, the higher Q is, the less is the permissible offset.

The deviation reduction factor of a process which is considered for proportional-position control can be determined approximately for a ¼-decay ratio from knowing its time constant and dead time, by means of the equation*

$$Q = \frac{t_2 + 0.8t_1}{t_2}$$

If the time constant of the above feedwater heater is 3 minutes and the dead time is 0.5 min., then

$$Q = \frac{0.5 + 0.8 \times 3}{0.5} = 5.8$$

* Based upon an article by Dr. W. Oppelt, "Rules of Thumb for Adjustments of Control Systems," (German), *Chemie-Ingenieur-Technik*, No. **8**, pp. 190–193, 1951.

which shows that the desired quality of control of $Q = 10$ would not be obtained. The feedwater temperature may drop to 176.55°F, since

$$\frac{180 - 160}{180 - 176.55} = 5.8$$

which corresponds with the deviation reduction factor of the particular process. Another way of improving Q is to increase the amplitude of the oscillations. This method shall be ignored since it generally is not acceptable.

If such a drop is not permissible, the proportional-position controller cannot be used. The deviation reduction factor is therefore a means to determine if a proportional-position controller is acceptable for a specific control system.

Adjustment of Proportional Band. The proportional band is the change of controlled variable necessary to move the final control element through its full stroke. It is usually, as it is here, expressed in per cent of the total measuring range of the controlling instrument. An instrument with total range from 0° to 200°F may have a 10 per cent proportional band. This means the final control element moves from fully open to fully closed for a temperature change of 20°F. The proportional band may also be expressed in units of the controlled variable that correspond to the full movement of the final control element. For example, a proportional band of 10°F means that the final control element moves through its full stroke when the temperature of the controlled variable changes by 10°F. For smaller temperature changes the movement is correspondingly smaller.

The narrower the proportional band, the faster the response, the smaller the offset, but the greater the tendency to oscillate. Figure 4–19

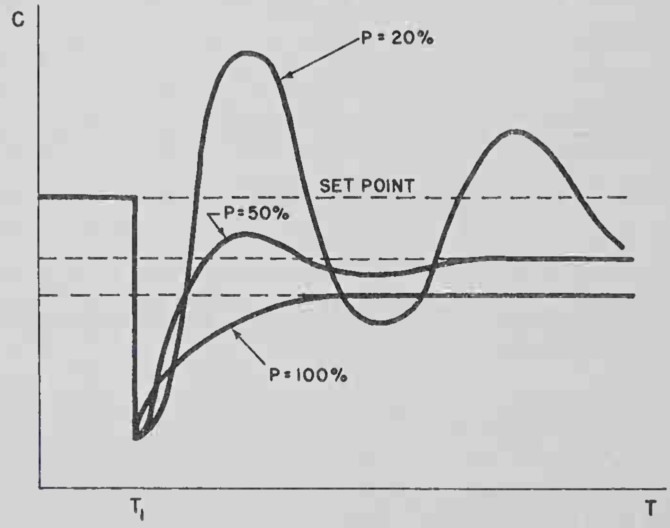


Figure 4–19. Controlling action with various proportional bands.

shows the typical response of a process with a proportional controller. The pH of a chemical solution may be considered the controlled variable. A sudden change of alkalinity occurs at time T_1. With a proportional band of $P = 20$ per cent, the controlled variable returns very quickly to the set point, but continues to oscillate around it. Since these oscillations subside gradually, they may be acceptable, particularly if a mixing tank provides equalization of the solution. Assuming however, that these oscillations are considered excessive, a proportional band of 50 per cent may be chosen. The response is slower, the overshooting is reduced considerably, but offset is the consequence. A further increase of the proportional band to 100 per cent results in an even larger offset but overshooting is now eliminated completely. An adjustment for a ¼-decay ratio would presumably lie between 25 and 30 per cent.

The proportional band of the controller for a ¼-decay ratio can be determined either by experiment or by calculation on the basis of certain numbers that have to be known. The method of calculation shall be discussed first.

Either one of the following equations may be used:

$$P = 100vt_2\frac{S}{R}$$

and

$$P = 100g\frac{t_2}{t_1}\frac{S}{R}$$

where the symbols have the following meaning:

P is the proportional band in per cent of the measuring range of the controlling instrument;
v is the process response rate (see Chapter 3) in units of the controlled variable/min./in. of motion of the final control element;
t_2 is the dead time in minutes;
S is the stroke in inches which the final control element moves when the controller output signal changes through its full range;
R is the measuring range of the controlling instrument in units of the controlled variable;
g is the process gain in units of the controlled variable/in. of stroke of the final control element;
t_1 is the time constant in minutes.

The magnitudes of time constant and dead time do not refer to the process alone, but to the results of a step response which includes final control element, process and measuring element.

For example, the maximum response speed of a process to a valve displacement of one inch is 20°F/min. The dead time is 10 min. The

stroke is 2 in. and the measuring range of the controlling instrument is 200°F. The resulting proportional band is

$$P = 100 \times 20 \times 10 \frac{2}{200} = 200 \text{ per cent}$$

The experimental method assumes that the control system is in operation and that a limited amount of process disturbances are permissible. The first step is to find the critical setting of the proportional band adjustment which means reducing the proportional band to the point where continuous oscillations of constant amplitude begin to appear. This must be done in successive steps. After each change of the proportional band, a load change is introduced by a small change of the set point, and the response is observed. If the response is still too stable, the proportional band must be further reduced.

The second step consists only in adjusting the proportional band to twice the critical setting. This is then the correct adjustment for ¼-decay ratio. For example, if the critical setting at which continuous oscillations begin to appear is $P = 10$ per cent, then the suitable proportional band for the process is $P = 20$ per cent.

Proportional-position vs. Proportional-speed Floating Action. The factor that limits the application of floating action is the floating rate. The stability of the control system may require a very slow floating rate. In this case, correction for a deviation can become so slow that the time which elapses before the controlled variable returns to the control point is longer than permissible for satisfactory control.

A slow floating rate may be compared with a wide proportional band of a proportional-position controller. In either case, speed of corrective action is limited in the interest of stability. From the equation

$$P = 100g \frac{t_2}{t_1} \frac{S}{R}$$

follows that low t_1/t_2 ratios (i.e. high t_2/t_1 ratios) diminish the response speed of proportional-position action.

The equation for proportional-speed floating control

$$f = \frac{k}{g(t_1 + t_2)}$$

shows that the t_1/t_2 ratio affects floating control only to the extent of the constant k. The table on p. 42 shows that the maximum effect of k does not exceed the ratio 5:1.

The magnitude of the sum of t_1 plus t_2, however, has a much more

pronounced effect on floating control than it would have on proportional-position action.

The conclusion is that proportional-position control is particularly valuable where large values of dead time and time constant are concerned, but where the ratio t_1/t_2 is also large. For proportional-speed floating control the reverse is true and processes with small values of dead time and time constant and small t_1/t_2 ratios are best suited.

For example, a temperature-controlled process with a time constant of 3 min. and a dead time of 1 min. is suitable for proportional-position control, disregarding offset. On the other hand, for a pH control system which may have a time constant of 0.02 min. and a dead time of 0.48 min., a proportional-speed floating controller would probably be the best answer.

Proportional plus Reset Action

Figure 4–20 illustrates proportional plus reset action in a pneumatic arrangement. The same mechanism as in Figure 4–17 is used for this

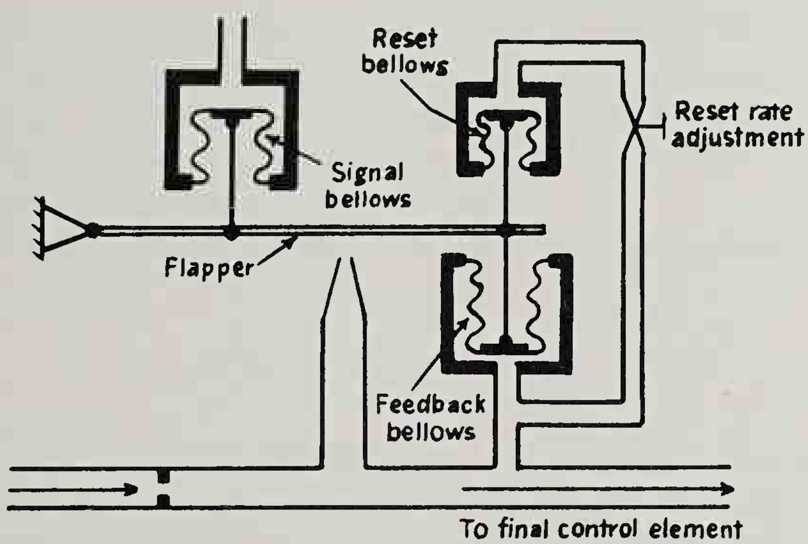


Figure 4–20. Mechanism showing proportional plus reset action.

example; the pilot valve, however, is ignored for the sake of simpler presentation. The reset bellows is added.

Suppose that initially the pressure in both the reset and the feedback bellows is equal and that the pressure in the signal bellows corresponds to the set point pressure. The torques exerted by the three bellows on the flapper balance each other. If the signal increases, the flapper moves clockwise, tending to close the nozzle. This results in an increase of nozzle back pressure which is also applied to the feedback bellows. The resulting position of the flapper, and hence the nozzle back pressure, is

thus determined by a new balance of forces exerted by the bellows. However, the increased nozzle back pressure gradually leaks through the restriction of a needle valve designated in the illustration as reset rate adjustment. In doing so the reset bellows continues to move the flapper in the clockwise direction which began with the action of the signal bellows. The nozzle back pressure continues to rise in an effort to increase the controller output signal to a point where the controlled variable finally returns to the set point.

Proportional plus reset action is a combination of proportional-position action and proportional-speed floating action. The result is elimination of offset from proportional-position action.

Reset or Integrating Action. Figure 4–21 compares the step function responses of the three actions: proportional-speed floating, propor-

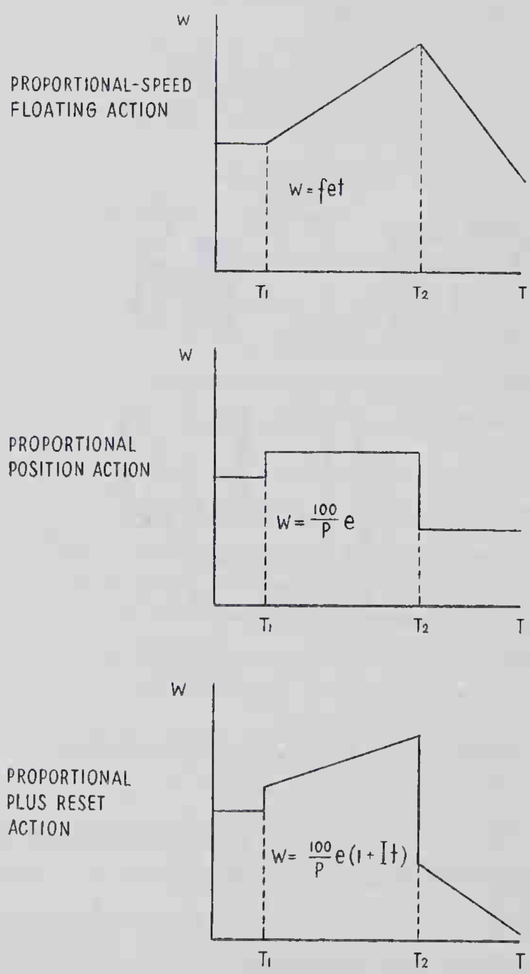


Figure 4–21. Step function response of different controllers.

tional-position, and proportional plus reset. Two step inputs have been applied as actuating signals at times T_1 and T_2 respectively. The second input was twice as large and in an opposite direction to the first one.

The action of the proportional-speed floating controller is expressed by

$$w = fet$$

where w is the change of controller output signal, f is the floating rate, e is the actuating signal and t is the time interval measured from T_1 and T_2, respectively. This means that as long as there exists an actuating signal, the controller output signal will continue to increase (or decrease) since the actuating signal e is multiplied by the time t. This continuous action which is a function of time is called *integrating action.* Offset cannot exist with integrating action, since an actuating signal will always result from offset. This is because offset is a deviation of the controlled variable from the set point, and the actuating signal is the difference between controlled variable and set point.

In the proportional-position controller, as shown in Figure 4–21, the action is expressed by

$$w = \frac{100}{P} e$$

where P is the proportional band of the controller. In this case the actuating signal e produces a change in controller output signal w of finite magnitude. If the change is not large enough, the result is offset, which results because the proportional-position controller does not provide integrating action.

The equation which corresponds with the proportional plus reset controller reads

$$w = \frac{100}{P} e + \frac{100}{P} eIt = \frac{100}{P} e(1 + It)$$

The first term contains the proportional-position control only. The second term corresponds with the floating controller, but the floating rate f is now replaced by $100(I/P)$. The factor I is the reset rate of the controller. It is multiplied by the elapsed time t. Due to this integrating action the controller eliminates offset.

The step function response for proportional plus reset action as illustrated in Figure 4–21 shows clearly the distinction between the initial sharp increase due to the proportional action and the integrating action of the reset. A proportional plus reset controller may well be considered a two-speed proportional-speed floating controller. First is the fast change, followed by a more gradual change. This becomes particularly obvious when the action of the final control element is considered; this follows only gradually a vertical change of the controller output signal. Thus the change from the action due to proportional control to that due

to reset action is only a change in speed in the response of the final control element.

The Reset Rate. The reset rate I in the above equation depends on the opening of the needle valve of the reset rate adjustment (see Figure 4–20). It also depends on the pressure differential across it and hence on the magnitude of change in controller output signal.

Figure 4–22 further illustrates the reset rate. In this case the initial corrective action of a controller, in response to a given deviation of the

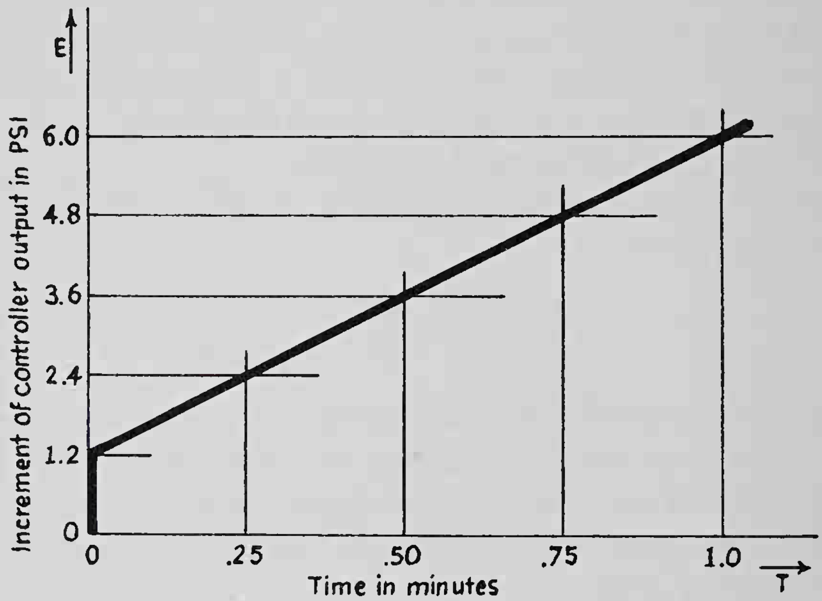


Figure 4–22. Proportional plus reset action.

controlled variable, is an increment of the transmitted pressure of 1.2 psi which is due to proportional-position action only. This corrective action is increased through the effect of reset action by 4.8 psi/min. Hence the reset rate I is equal to 4 repeats/min.

Suppose this controller has the usual output of 3 to 15 psi which is the signal range to the final control element. The corrective action at $T = 0$ is due to proportional-position action only. It is 1.2 psi or 10 per cent of the output range ($15 - 3 = 12$ psi). Hence $w = 10$ per cent at $T = 0$. If the proportional band $P = 50$ per cent, then at $T = 0$

$$e = \frac{100}{Pw} \quad \text{or} \quad e = 5 \text{ per cent}$$

which means that the response as illustrated was caused by a deviation of the controlled variable from the set point of 5 per cent of the instrument range. Since the reset rate I was stated to be 4 repeats/min. it follows that after 1 min.

$$w = \frac{100}{50} 5(1 + 4) = 50 \text{ per cent}$$

which expressed in psi of controller output change is 6 psi, a figure which corresponds with Figure 4–22.

Adjustments of Proportional Band and Reset Rate. The adjustments of the proportional plus reset controller may again be made either by experiment or by calculations. Given the measuring range R of the controller and the stroke S of the final control element, and knowing, furthermore, either the process response rate and the dead time, or the process gain, time constant, and dead time, the adjustments can be made by following the expressions

$$P = 100vt_2\frac{S}{R}$$

or

$$P = 100g\frac{t_2S}{t_1R}$$

These equations are the same as those used before for the proportional-position controller. The reset rate is given by

$$I = \frac{0.3}{t_2}$$

The example which was used for the proportional-position controller had a dead time of 10 minutes. Adding reset to the controller would require a reset rate of

$$I = \frac{0.3}{10} = 0.03 \text{ repeats/min.}$$

The experimental method follows the steps outlined for proportional-position control. In determining the critical setting of the proportional band at which continuous oscillations begin to appear, the reset action is cut out. The time in minutes, T, between two successive maxima of oscillations at the critical setting is measured. The suitable final adjustment of the proportional band is the same as for proportional action alone, i.e, twice the critical setting, and the reset rate is

$$I = \frac{1.2}{T}$$

Proportional plus Reset plus Rate Action

The proportional band of a controller can only be reduced to a certain point. Beyond this point the control system becomes unstable. This limits the speed of correction. The wider the band, the less the sensitivity and hence the slower the correction. This lag in correction becomes particularly noticeable where the time constant of the measuring element is of noticeable magnitude.

A typical example is a multi-capacitance temperature-controlled system. In this case the equivalent dead time is relatively large because of the multiple capacitances. This requires a wide proportional band. In addition, the measuring lag of the primary element is considerable. This results in relatively slow corrections for the deviations of the controlled variable from the set point.

It is desirable in such cases to create a response to a slow change of the actuating signal by an acceleration of this signal in the controller. This gives faster response, but one which slows down as the controlled variable returns to the set point, thus preventing excessive overshooting and hence instability. Controllers can be provided with rate action which results in the described response. Occasionally, proportional action may be provided with rate action only. However, this is rather rare. More frequently the controller in this case includes also reset action and becomes a proportional plus reset plus rate controller. The reason is that a proportional controller that can operate with a narrow proportional band does not generally require reset action, since the offset is insignificant. In this case, however, the narrow band provides quick enough corrective response and rate action is hardly warranted. As the band becomes wider, the response is slower, the offset larger, and reset action becomes necessary with rate action added for certain conditions.

Figure 4–23 is a schematic illustration of proportional plus reset plus rate action. It differs from Figure 4–20 in that a fourth bellows—the rate bellows—is added, and that the area of the feedback bellows is smaller. When a change in signal occurs, the deflection of the flapper is much larger, due to the reduced area of the feedback bellows. The controller output signal as transmitted to the final control element is assumed

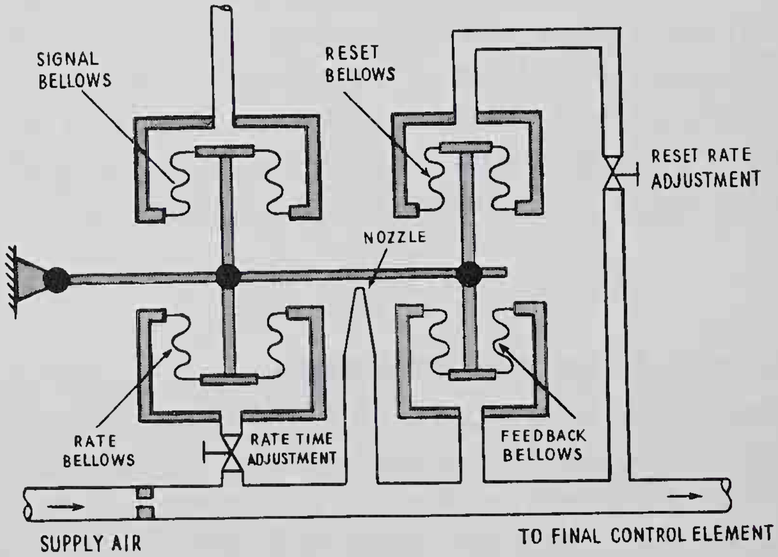


Figure 4–23. Proportional plus reset plus rate action.

to be proportional to the flapper deflection. A change in controller output signal is gradually transmitted to the rate bellows through the restriction of the rate time adjustment. This reduces the initial deflection of the flapper.

If the pressure of the controlled variable, as applied to the signal bellows, changes slowly, the pressure of the controller output signal will also change slowly and there will be no large difference between the pressures in the feedback and rate bellows. However, the faster the controlled variable changes, the larger will be this pressure difference and the greater will be the effect of the rate action.

The response of the proportional plus reset plus rate controller to a step input would be maximum and only limited by the mechanical conditions of the system. Hence, the inherent characteristics could not be fully represented by a step function response. The change of the controlled variable is therefore represented by a ramp function, i.e., an input signal of a more gradual changing magnitude as illustrated in Figure 4–24. At T_1 the input signal U begins to decrease. The controller

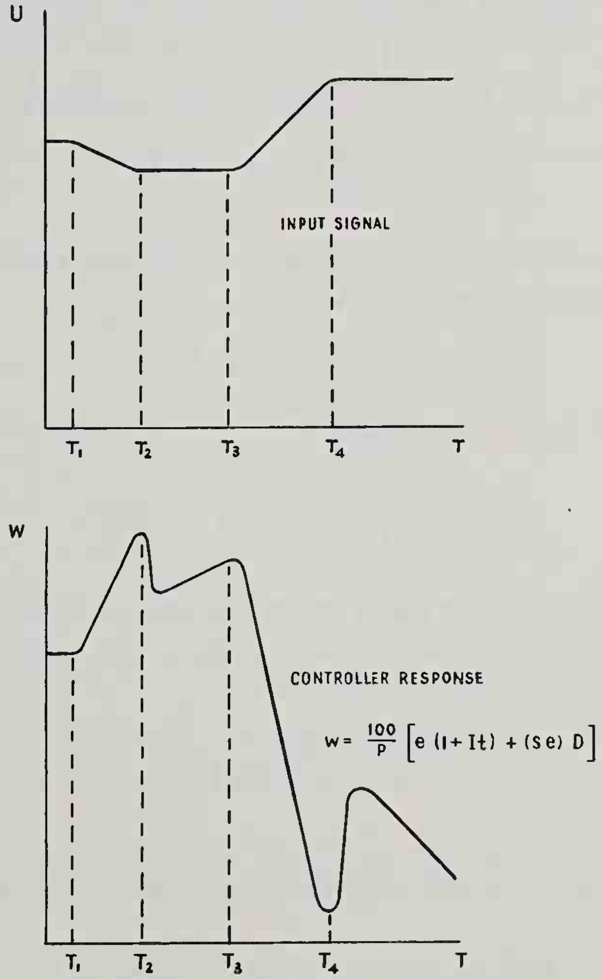


Figure 4–24. Response of proportional plus reset plus rate controller.

output signal rises correspondingly but at a faster rate. At T_2 the input signal stops decreasing and remains constant at a new level. The effect of the rate action subsides immediately to a point which is determined by the proportional band. However, the action of the reset mechanism also enters and continues to increase the controller output signal, though at a slower rate. From time T_3 to T_4, the input signal increases but this time at a faster rate, and the magnitude of the change is also larger. The action is correspondingly reflected in the controller output signal.

Rate Time. The mathematical expression for proportional plus reset plus rate action is

$$w = \frac{100}{P}e + \frac{100}{P}eIt + \frac{100}{P}(se)D = \frac{100}{P}[e(1 + It) + (se)D]$$

The last term stands for the rate action. The rate of change of the input signal e is expressed by (se). For example, if the input signal at a given moment changes 3 psi/min., then $(se) = 3$. The constant D expresses the adjustable rate time of the controller. Rate time is generally given in minutes and indicates the time difference in response between a proportional and a proportional plus rate controller. For example, assume that the controller has no reset action, i.e., $I = 0$, then

$$w = \frac{100}{P}[e + (se)D] \tag{1}$$

Let the controlled variable change at a certain rate (se), then after a time T_a, the controlled variable will be

$$e = (se)T_a$$

Equation (1) can then be written

$$w = \frac{100}{P}[(se)T_a + (se)D] \tag{2}$$

Suppose the rate action to have been cut off, then it would have taken a longer time—expressed by T_b—to obtain the same magnitude w of controller output signal, or

$$w = \frac{100}{P}(se)T_b \tag{3}$$

Equations (2) and (3) both express w of the same magnitude, hence

$$\frac{100}{P}[(se)T_a + (se)D] = \frac{100}{P}(se)T_b$$

and this reduces to

$$D = T_b - T_a$$

which shows that the rate time expresses the difference between time due to proportional action only and the time that it takes for the combination of proportional plus rate action to obtain the same controller output.

For example, if P, the proportional band, is 25 per cent of the measuring range of the controller and if the controlled variable changes at the rate of 50 per cent/min., i.e., $(se) = 50$, then with proportional action only, the controller output after 1 min. is

$$w = \frac{100}{50} 50 = 100 \text{ per cent of its range.}$$

Adding rate action with a rate time of 0.25 min., the same change should be obtained in $1 - 0.25 = 0.75$ min. This can be checked by inserting values into equation (2), which gives

$$w = \frac{100}{50}(50 \times 0.75 + 0.25 \times 50) = 100 \text{ per cent}$$

Adjustments of Proportional Band, Reset Rate, and Rate Time. The equations which can be used to determine adjustments are: for the proportional band;

$$P = 80\, v t_2 \frac{S}{R}$$

or

$$P = 80 g \frac{t_2}{t_1} \frac{S}{R}$$

for the reset rate;

$$I = \frac{0.5}{t_2}$$

and for the rate time;

$$D = 0.5\, t_2$$

using the same example as for proportional-position action, the proportional band is

$$P = 80 \times 20 \times 10 \frac{2}{200} = 160 \text{ per cent,}$$

the reset rate is

$$I = \frac{0.5}{10} = 0.05 \text{ repeats/min.,}$$

and the rate time is

$$D = 0.5 \times 10 = 5 \text{ min.}$$

The experimental method proceeds along the steps outlined for proportional plus reset controllers. The adjustments are then as follows:

1. The proportional band should be 1.7 times the critical setting.
2. The reset rate should be

$$I = \frac{2}{T}$$

   where T is the time between two successive maxima of oscillations at the critical setting.
3. The rate time should be

$$D = \frac{T}{8}$$

5. FREQUENCY RESPONSE

Frequency response methods are more cumbersome than step function methods because they require special equipment and are more time consuming. From the practical viewpoint, they are justified because they allow division of a control loop into an arbitrary number of elements, to obtain frequency response curves of each of the elements, and to predict their behavior in the loop or in combination with any other components of which the frequency response is known.

Frequency response is similar to step function response in that it compares the output of a component or of a group of components with the input. It is similar also in its object which is to obtain a stable control system with maximum correction speed. The difference between the two is the form of the input signal, which is a sudden change in one case, and a continuously changing signal in the other.

Sinusoidal Input

The pattern which the continuously changing signal follows is that of a sinusoidal curve, such as shown in Figure 5–1. The reason for choosing

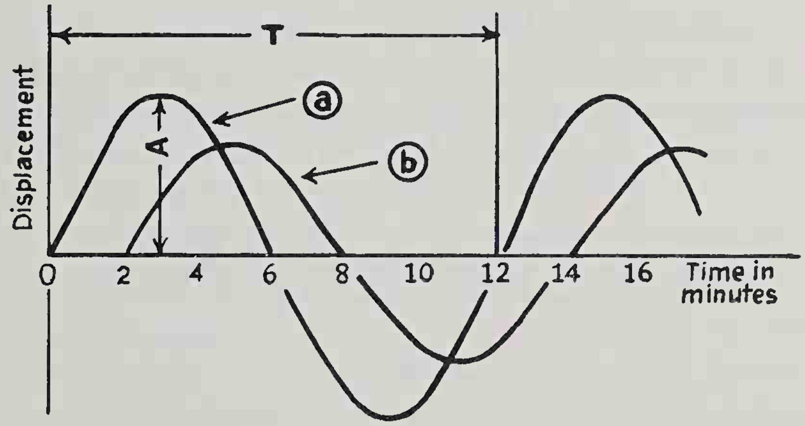


Figure 5–1. Sinusoidal curves.

such a signal is mainly to allow correlation of mathematical and experimental results, which however is of more academic than practical interest.

There is, however, a deeper reason for choosing the continuously changing signal, in that many physical systems have responses of a simple sinusoidal pattern. Wherever an elastic restoring force is involved, e.g.,

in the springs of a car, the pendulum of a clock, the mercury column in a manometer, etc., the system will tend to restore its balanced conditions by sinusoidal motions once its equilibrium is affected by an outside force. A periodic motion of sinusoidal pattern, repeating itself at definite intervals of time, is called simple harmonic motion. The amplitude of a sinusoidal curve is defined as the distance from the mean value to the point of maximum displacement. This is indicated by A in Figure 5–1, curve *a.* The time required for the motion to complete one cycle is called the period. Thus, in Figure 5–1, for either curve *a* or *b,* the period T is 12 min.; after this the pattern repeats itself.

Once the period and the amplitude are known, it is possible to draw a sinusoidal curve with the help of a trigonometric table, or simply by using the following values:

Angle	Sine
0°	0.000
15°	0.259
30°	0.500
45°	0.707
60°	0.866
75°	0.966
90°	1.000

The period T is subdivided into 360 degrees. In Figure 5–1 this interval coincides with 12 min. Hence each minute is equal to 30 degrees. The amplitude is multiplied by the sine value of various points to obtain the displacement at these particular points. Thus the following table is established for curve *a* of Figure 5–1, which has an amplitude equal to A:

Time in Minutes	Angle in Degrees	Displacement
0	0	0.000 A
0.5	15	0.259 A
1.0	30	0.500 A
1.5	45	0.707 A
2.0	60	0.866 A
2.5	75	0.966 A
3.0	90	1.000 A

Thus enough points can be established to draw a curve up to 90 degrees. From there the preceding table can be used up to 180 degrees

by taking the values in descending order, i.e., 0.966 *A*, 0.866 *A*, 0.707 *A*, etc., for successive intervals of 0.5 min. It is obvious that by using negative values the curve can be completed.

Phase Lag, Frequency, and Magnitude Ratio

Figure 5–1 shows two curves—curve *a* and curve *b*. These can be visualized as the input and the output of a process, e.g., in a gas-heated steam generator.* Suppose the gas fuel control valve were to be cycled according to curve *a*, then the steam pressure would follow with a similar pattern, only delayed in action. This delay is equal to 2 minutes in Figure 5–1. Such a delay in sinusoidal response is called a *phase lag* or *phase angle*, and is expressed in degrees. As previously established, the period *T*, 12 min. in the case under consideration, corresponds to 360 degrees, and hence the phase angle is -60 degrees, where the minus sign indicates that the output lags behind the input.

A period of 12 min. means that a complete cycle of the periodic motion is completed in this time interval. Conversely, this means that $1/12$ of the cycle is completed in 1 min. The frequency of a sinusoidal input is expressed in those terms, i.e., $1/12$ cycle/min. is the frequency of either curve in Figure 5–1. Thus, the *frequency* is the inverse of the period.

The higher the frequency of the fuel valve, which means the faster the valve is moved through its periodic motion, the more difficult it becomes for the pressure to follow the input. Although the frequency of both input and output is always the same, the phase angle will increase and the amplitude of the pressure curve will get smaller. Finally a frequency is reached which is too fast for the pressure to follow and it will stay constant at some average value. The ratio between amplitude of output signal to the amplitude of input signal is called the *magnitude ratio.*

Technique of Taking Frequency Response Data

Suppose it is desirable to take the frequency response of a steam generator control system. The controller is assumed to be air operated and with proportional-position action. Fuel is admitted to the generator, which produces steam. The pressure of the steam is measured and the controller puts out a signal as a function of the deviation of the steam pressure from the set point. The control signal is applied to a control valve which regulates the fuel flow. The interaction of the closed loop is thus established.

---

* Linear relations are assumed throughout this chapter. This is generally permissible. As pointed out, however, at the end of this chapter, the interpretation of frequency response data has its limitations largely because of this assumption.

In order to obtain the frequency response of such a system it is necessary to open the loop at some arbitrary point, e.g., between controller and valve, as shown in Figure 5–2. The closed loop thus becomes an open loop.

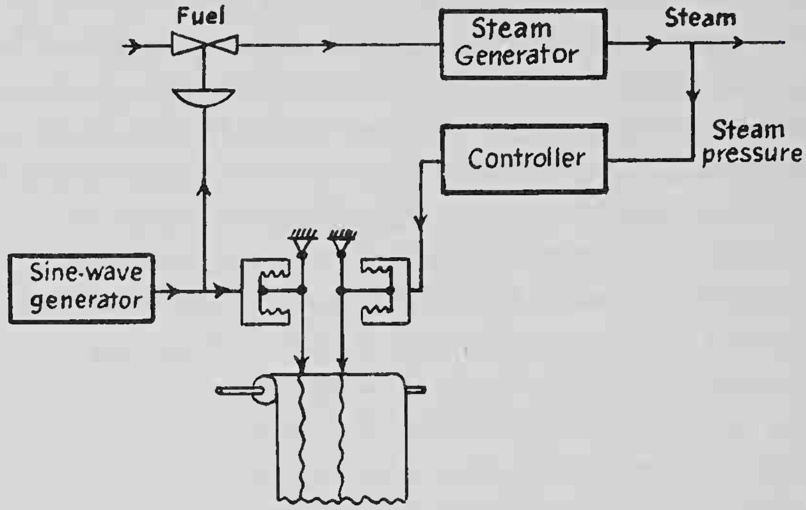


Figure 5–2. Open loop frequency response.

An input signal is applied to the control valve. This requires a pneumatic sine-wave generator, which contains a variable-speed electrical motor. The rotation of the motor shaft is converted into simple harmonic motion by means of a cam, linkage, or slider. A pneumatic relay is operated by this linkage, thus providing an air pressure that cycles sinusoidally. By increasing the supply pressure or varying lever ratios, the amplitude of the signal can be changed. The frequency is regulated by the motor-speed control. The arrangement is basically the same for hydraulic and mechanical sine-wave generators. Where electrical signals can be used, a number of sine-wave generators are commercially available.

Another piece of equipment is the recorder illustrated in Figure 5–2, which records graphically both the input and the output of the open loop. It is important that this recorder does not add to the attenuation and phase shift of the input signal; i.e., the frequency response of the testing equipment must be considerably above the frequency applied in the test of the open loop. Frequently, oscilloscopes and other electronic equipment are used for this purpose.

A certain amplitude is chosen for the frequency test. For reference purposes, it is quite necessary to always note the amplitude at which response data were taken. The amplitude is limited by the speed of the components. For example, a valve may move 1 in. in 5 sec. Choosing an amplitude which corresponds to a valve motion of one inch limits the

frequency to less than 3 cycles/sec. Thus the speed and not the frequency may become the limiting factor.

The difference between speed and frequency response is an important one. A system may be rather slow, but follow the input signal without attenuation. With a small enough amplitude, this system will respond satisfactorily even at high frequency. For larger amplitudes, however, the slowness of the system becomes the controlling factor. Whether speed or high frequency response or both are required depends entirely on the application.

Once the amplitude is chosen, a number of runs are taken, changing only the frequency of the input signal. It is assumed in the following, unless otherwise stated, that the proportional band of the controller is such that at low enough frequencies the magnitude ratio is 1. The frequencies are then increased, the data logged and put into graphic form.

Graphic Representation of Frequency Response Data

The frequency response of a system coordinates three variables: frequency, magnitude ratio, and phase lag. Figure 5–1 illustrated magnitude ratio and phase lag at one specific frequency, namely at $1/12$ cycle/min. This information does not suffice to draw any conclusions about the behaviour of the control system. For that purpose, it is necessary to graphically represent the change of magnitude ratio and phase lag at various frequencies. Several methods exist. The Nyquist diagram has been used for many years. It has the advantage of combining all the information in a single curve, but it is generally considered more difficult to interpret than the so-called Bode diagram, a method which conveys the same information as the Nyquist diagram, but in two separate diagrams. The American Society of Mechanical Engineers has accepted the Bode diagram as the preferred standard for the presentation of frequency response data.

Nyquist Diagrams. Let the control valve in Figure 5–2 have a total stroke of 2 in. The controller output signal has a range of 12 psi. Cycling the valve through an amplitude of 0.2 in. corresponds then to cycling it through an amplitude of 1.2 psi. Suppose at a certain frequency the controller output signal responds to this input with a maximum amplitude of 0.96 psi, then the magnitude ratio of output to input is $0.96/1.2 = 0.8$.

Figure 5–3 represents a Nyquist diagram corresponding to the control system of Figure 5–2. At 0.01 cycle/min. the magnitude ratio between output and input is 1.0. The length of the arrow or vector *a* in Figure 5–3 is proportional to the magnitude ratio. The phase angle is zero.

At 0.03 cycle/min., the phase angle is −30 degrees, although the magnitude ratio remains the same. The vector *b* is hence equal in length to vector *a,* but it is displaced by −30 degrees as indicated in Figure 5–3.

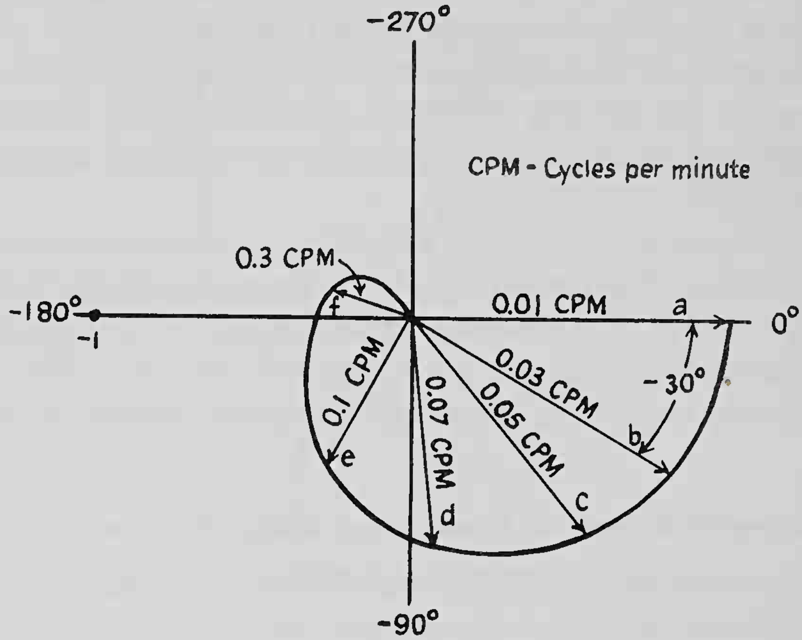


Figure 5–3. Nyquist diagram.

At 0.05 cycle/min., the phase angle is −50 degrees and the magnitude ratio is 0.9, which is expressed by making vector *c* correspondingly shorter.

Vectors, *d, e,* and *f* are of increasing negative phase angle and decreasing magnitude ratio. Actually at 0.7 cycle/min. the magnitude ratio decreases to 0.16 psi/ in. which is too small a value to enter on the plot.

By connecting the end points of the vectors, a curve is obtained which can be used for the interpretation of the control system. The Nyquist criterion states that *a system will be stable as long as the −1 point* (on the horizontal axis of Figure 5–3) *is to the left of the curve.*

Figure 5–4 shows several curves obtained with proportional–position control systems. Curve 1 is typical for a single capacitance system and curve 2 for a two-capacitance system. Curve 3 represents a three-capacitance system. These three systems are stable, although curve 3 gets close to the −1 point, which indicates that some oscillations will be unavoidable. In the case of curves 4 and 5, however, the systems—also of three capacitances—will be definitely unstable since the −1 point is no longer to the left of the curve.

The number of capacitances in a proportional-position control system can be read from the Nyquist diagram by dividing it into four quadrants, as indicated by numbers I, II, III and IV in Figure 5–4. Curve 1 remains

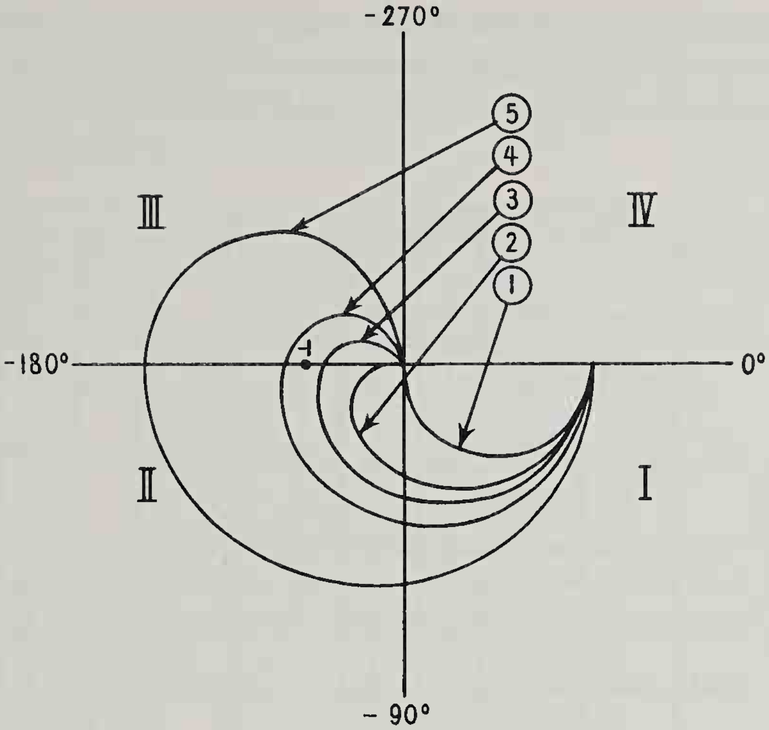


Figure 5–4. Nyquist diagrams for various processes.

in the first quadrant, hence it represents a single capacitance system, curve 2 enters the second quadrant indicating a two capacitance system, and curves 3, 4, and 5 stand for a three capacitance system since they enter into the third quadrant.

The Bode Diagram. Figure 5–5 gives the information of Figure 5–3, but in the form of a Bode diagram. The upper curve illustrates the change of magnitude ratio with increasing frequency. The scales used are not linear, but are so-called logarithmic scales. The distance from 0.01 to 0.1 is the same as the distance from 0.1 to 1.0. Similarly, the steps from 0.2 to 0.4 to 0.6 to 0.8 get progressively shorter. Thus the scale decreases with increasing values. This has the advantage that the lower values are easier to read than the higher ones. It has the further advantage of replacing the multiplication of frequency response data by graphic addition, as will be discussed later.

In the lower diagram, the phase angle is shown on a linear scale while the logarithmic scale continues to be used for the frequencies.

Phase Lag and Phase Reversal

Figure 5–2 illustrated the method of taking two simultaneous graphs in a frequency response test. As long as the output does not lag behind the input, the corresponding graphs would appear as in Figure 5–6, where one is the mirror image of the other. While the amplitudes of input and output signals are equal, the output is at maximum when the input

is at a minimum, and vice versa. Since a full period corresponds to 360 degrees, the distance between maximum and minimum of a sinusoidal curve is 180 degrees, and the output is said to lag behind the input by 180 degrees.

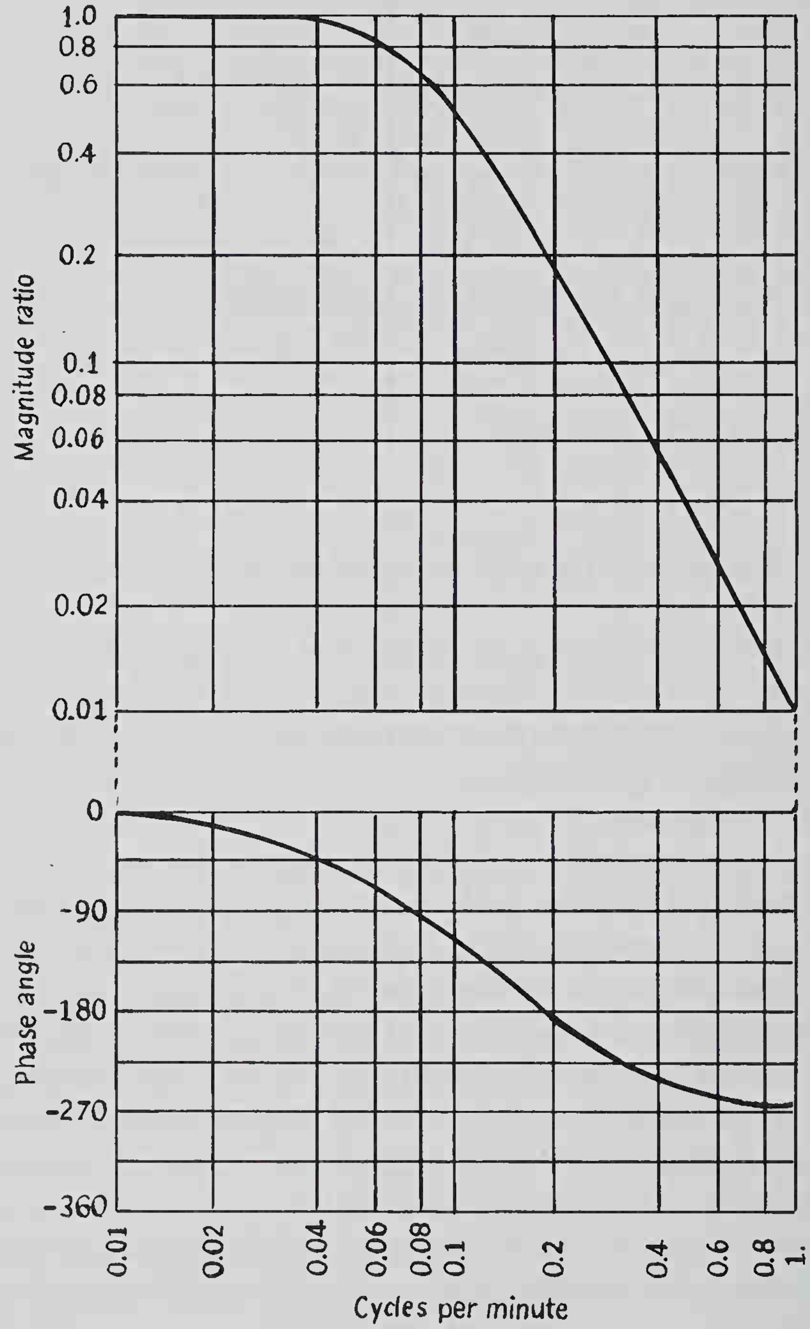


Figure 5–5. Bode diagram.

Actually this is not a phase lag but a *phase reversal* which is inherent in controller action. The controller has to counteract any disturbance in the loop. For example, an *increase* in steam pressure must result in such action of the controller as to *reduce* the steam pressure. This is equiva-

lent to saying that a 180 degree phase reversal as illustrated in Figure 5–6 is required. For the graphic representation of the frequency response, the phase reversal is ignored and only the additional phase shift is represented as phase lag.

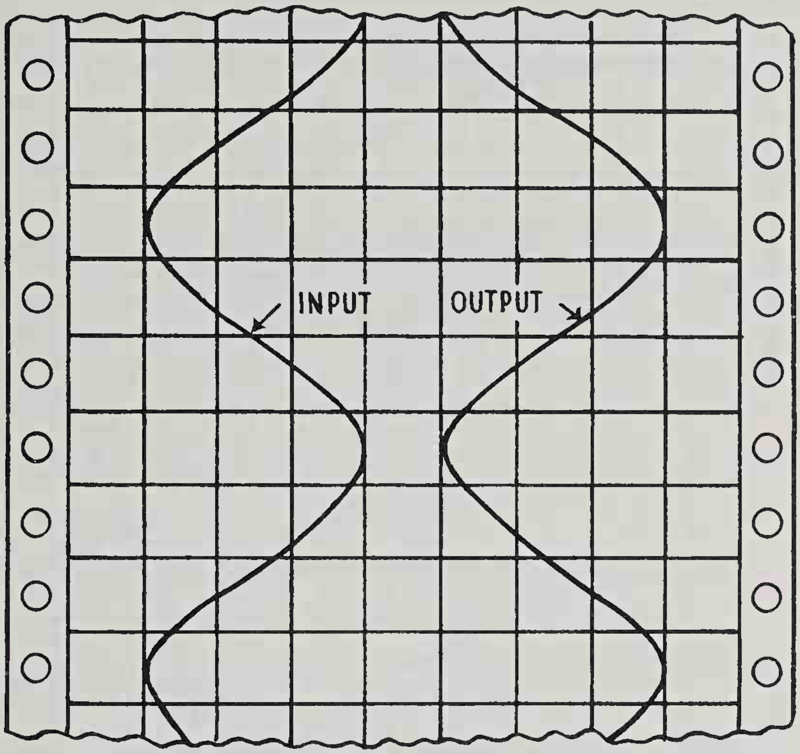


Figure 5–6. Response curves.

Stability Conditions

Suppose there is an additional phase shift of 180 degrees while the magnitude ratio is still unity. In this case the maxima of the output coincide with the maxima of the input. The controller actually repeats the disturbance, sustaining instead of counteracting it. The system is unstable.

The condition of unity magnitude ratio and 180 degree phase lag must therefore be avoided. It is common practice to accept safety margins according to the following rules:

1. The phase lag should not be more than 150 degrees when the magnitude ratio is one or more. The 30 degree difference between the acceptable and the unstable condition is called the *phase margin.*

2. At a 180 degree phase lag the magnitude ratio should be equal or less than 0.5. For a magnitude ratio of 0.5, it would be necessary to increase it by a factor of two in order to make it unity and hence make the system unstable. The factor by which the magnitude ratio has to be increased to obtain instability is called the *gain margin.* A gain margin of 2 is therefore desirable for process control.

Acting according to these stability criteria results in obtaining the fastest practical control response under stable conditions. Figure 5–7 shows frequency response curves which fulfill these conditions. The phase margin is 30 degrees and the gain margin is 2.

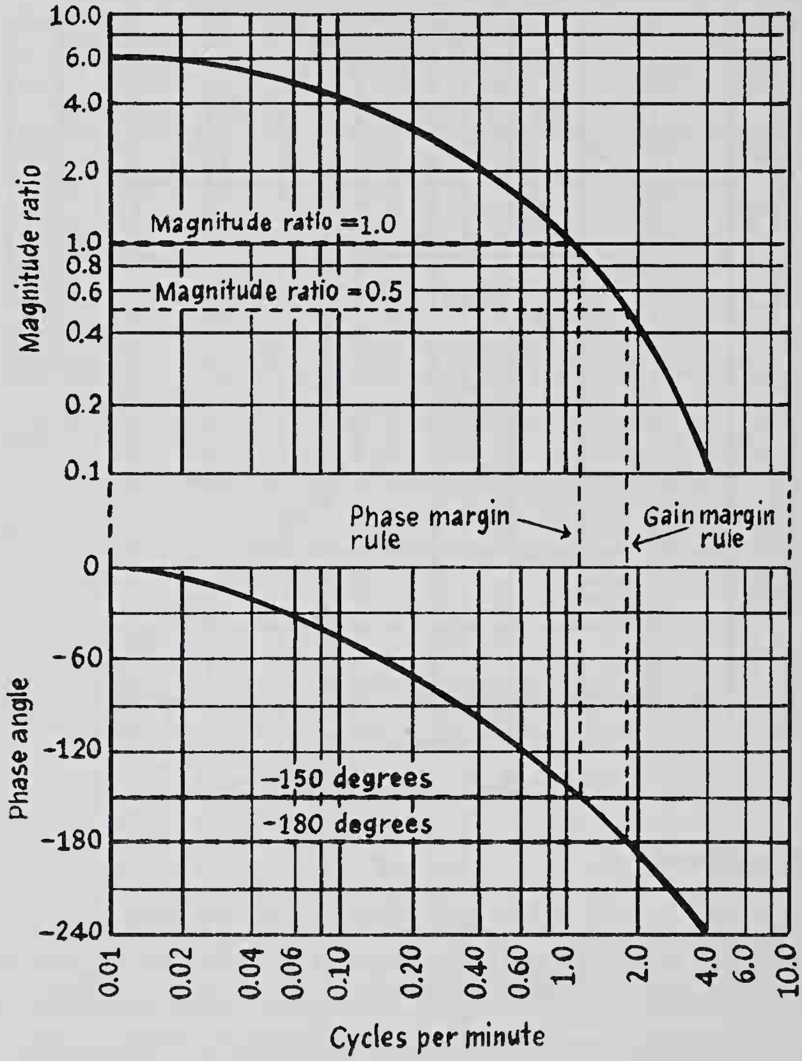


Figure 5–7. Frequency response with minimum conditions for stable system.

Dead Time as Instability Factor

It can be concluded from the foregoing that an increase of phase angle without a simultaneous decrease of magnitude ratio will lead to instability, excepting the case where the magnitude ratio is sufficiently below unity at all times. This condition, i.e., increase of phase angle at constant magnitude, is the result of dead time, since it represents the time it takes before a change in the loop becomes noticeable. Figure 5–8 illustrates the frequency response as it results from dead time in the system. This response indicates instability.

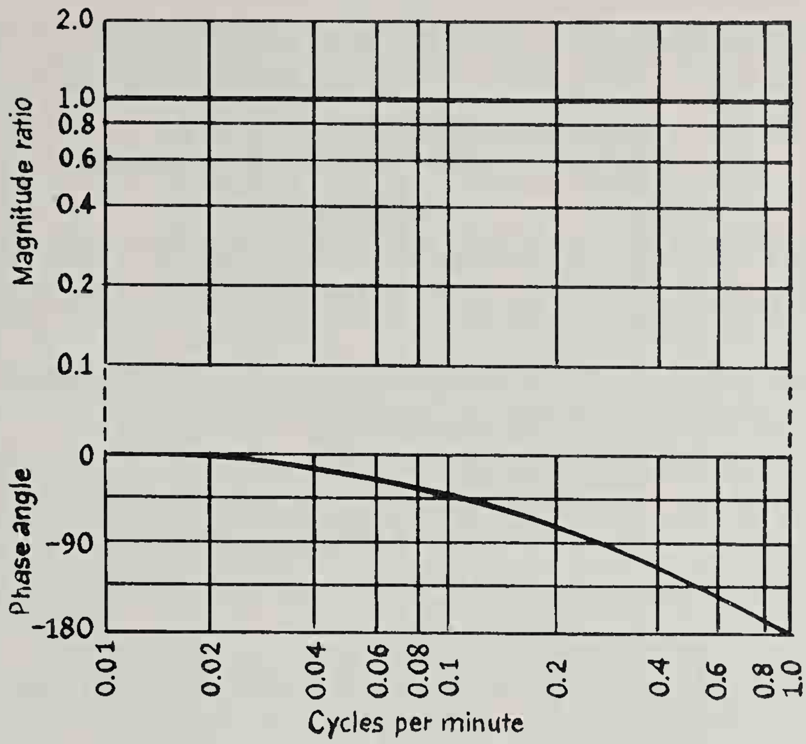


Figure 5–8. Frequency response of dead time.

Combination of Frequency Response Data

The frequency response of a controller or any particular component is of interest only if the frequency responses of all other components in the control loop are equally known. It is then possible to combine these data and obtain a complete interpretation of the behavior of the control system.

Block Diagram

Block diagrams illustrate how the frequency responses of various components in a control loop can be combined. Figure 5–9 shows a typical application of this method for a pressure control system. The pressure is applied to a diaphragm-and-spring combination. When, for example, the pressure increases, the jet pipe is deflected to the right. The deflection is proportional to the process pressure change. Oil is pumped through the jet pipe and leaves at its tip to be directed into two orifices which connect the two sides of an actuating piston. As the jet pipe is deflected toward the right, more oil passes into the right-hand than into the left-hand orifice. The piston moves toward the left, opening the control valve, thereby reducing the pressure in the process line.

In the block diagram, the control system is divided into its most obvious parts. This is quite arbitrary and different breakdowns are pos-

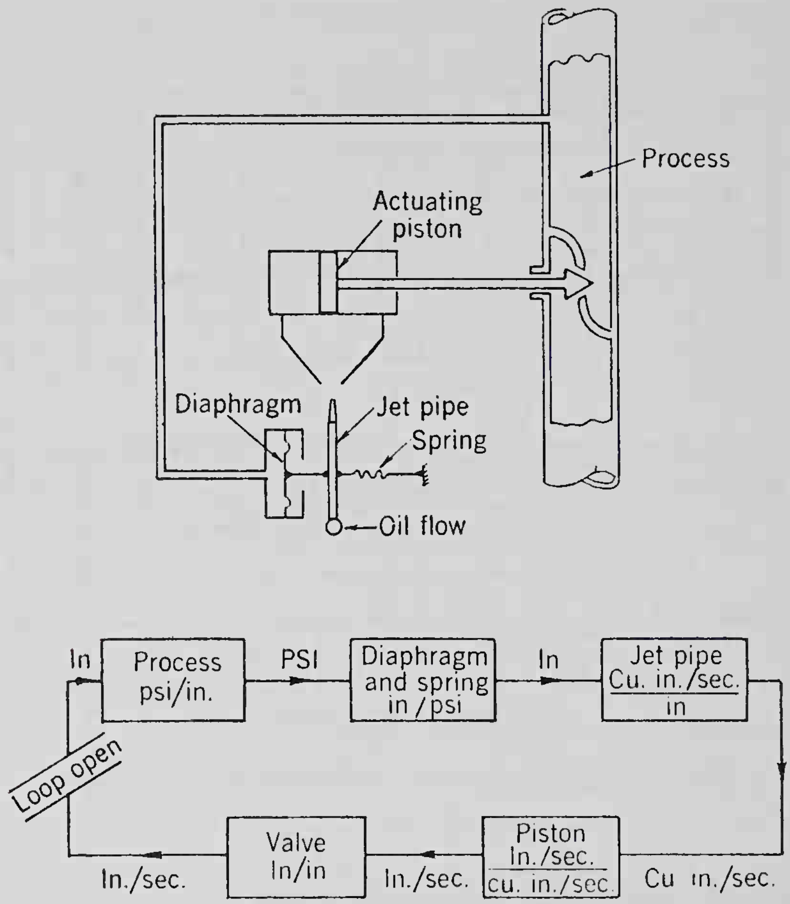


Figure 5–9. Block diagram of control system.

sible. The system is now represented by five blocks, namely; process, diaphragm-and-spring combination, jet pipe, piston, and valve.

The process changes its pressure per inch of valve motion (linear relations are assumed throughout). Hence the "process" block reads "psi/in." The valve motion expressed in inches is shown as input and the pressure in psi as output. This pressure is directed toward the diaphragm-and-spring combination, which produces a deflection of in./psi pressure change. This is converted into cu. in./sec. of hydraulic flow by means of the jet pipe, and the piston transforms the hydraulic flow into in./sec. of piston motion. The valve moves with the piston, and the final outcome is in./sec.

The important point is that a chain of multiplications has been performed by each block around the loop, namely

Process: in. × psi/in. = psi

Diaphragm-and-Spring: psi × in./psi = in., etc.

Each block acts as a multiplier upon its input. Each input is the output of another block. It is possible to lump together two or more of the blocks into one by a simple multiplication. For example, the "diaphragm-

and-spring" block can be combined with the "jet pipe" block by giving the new block the dimensions of:

$$\frac{\text{in.}}{\text{psi}} \times \frac{\text{cu. in./sec.}}{\text{in.}} = \frac{\text{cu. in./sec.}}{\text{psi}}$$

which is obviously in line with the input of psi and the output of cu. in./sec., since the multiplying of the new block with the input results in the output.

Each of these blocks could be separately submitted to a frequency response test. For example, the process could be analyzed by cycling the valve and measuring the pressure. The relation of output to input both in magnitude ratio and phase lag could thus be obtained. Similar data could be obtained from each block.

The output/input ratios (the magnitude ratios) can be multiplied to obtain the overall magnitude ratio, since they represent exactly the relationship represented in the block diagram. The phase angles, however, must be added since they represent delays in each block which simply add up as a signal passes through the loop.

Graphic Addition of Frequency Response Data

The magnitude ratios of frequency response data are represented in logarithmic scales, which resemble slide rule scales where the adding of two distances calibrated for the magnitude of the factors is equivalent to multiplication. The result is that the Bode diagram allows the graphic addition of magnitude ratios.

Since phase angles are represented on linear scales, their graphic addition corresponds with the physical addition required for phase angles.

The addition of magnitude ratios and phase angles is illustrated in Figure 5–10, where the frequency response data of two components *A* and *B* are combined and the result of this combination corresponds with *C*.

In order to add graphically magnitude ratio *B* to *A*, the ratio of 1 is used as a reference line. This is obvious when one considers that if the output equals the input, i.e., the ratio equals 1, then no matter how many components are connected together, their output to input ratio will still equal 1. It is also obvious because multiplication by 1 does not change the multiplicand.

The illustration shows three different arrows, *a, d,* and *g* which have been added to *b, e,* and *h,* respectively. For example, arrows *a* and *b,* which are taken at the same frequency, are of equal length but opposite polarity. Hence they cancel each other, which is expressed in *C,* by point *c* being located at unity magnitude ratio. Adding *d* and *e,* the opposite polarity has again to be taken into account. Therefore, the resultant *f*

corresponds to e diminished by d. The arrow i is simply the sum of g and h, both being of equal polarity. This method of addition results in the magnitude ratio C, which corresponds with the combined characteristics of components A and B. The addition of phase angles shown in Figure 5–10 uses a 0-degree phase angle as reference. When A lags by

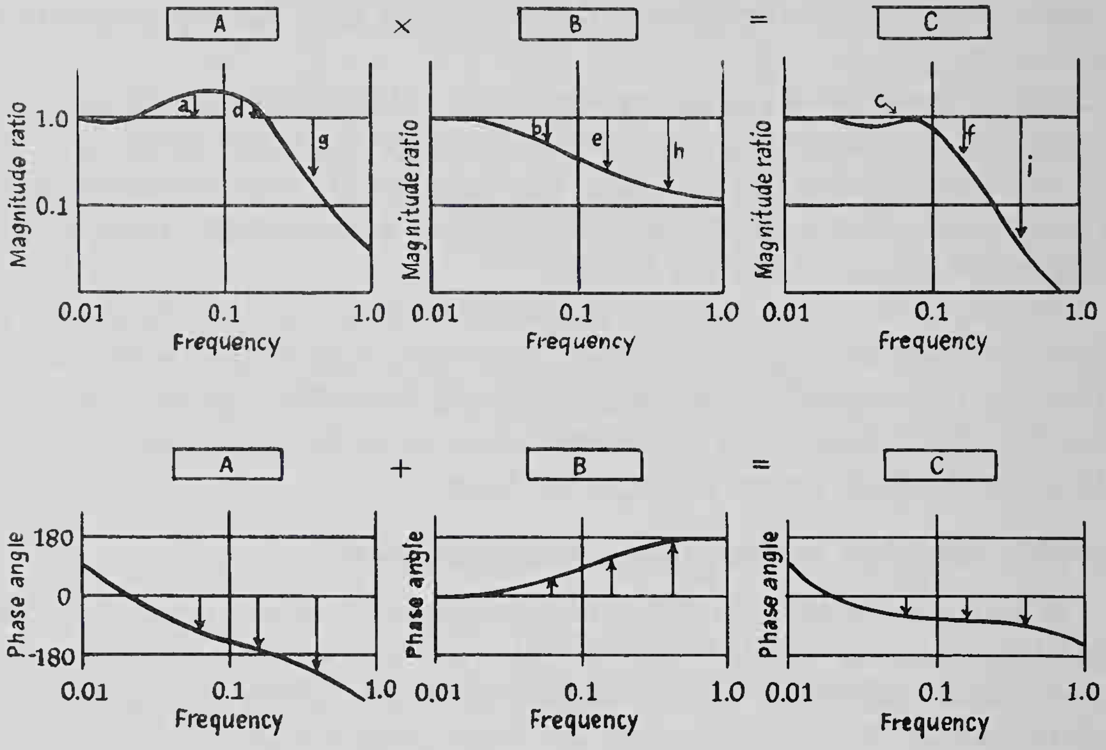


Figure 5–10. Addition of magnitude ratios and phase angles.

110 degrees at a certain frequency and B leads by 60 degrees at the same frequency, then the combined phase angle is -50 degrees. The illustrated addition is based on this method.

Unstable Components in a Stable Loop

It is interesting to observe that component A in Figure 5–10 is unstable, since the phase angle is -180 degrees at a frequency at which the magnitude ratio is equal to unity. By adding B to A, the result is C, and the combined response shows stability. This proves that a stable control loop may contain components which, considered by themselves, are unstable.

Proportional-Speed Floating Controller

Figure 5–11 shows the response of a proportional-speed floating controller. As long as an input signal is applied, the output of such a controller continues changing at a certain speed. When the change of the

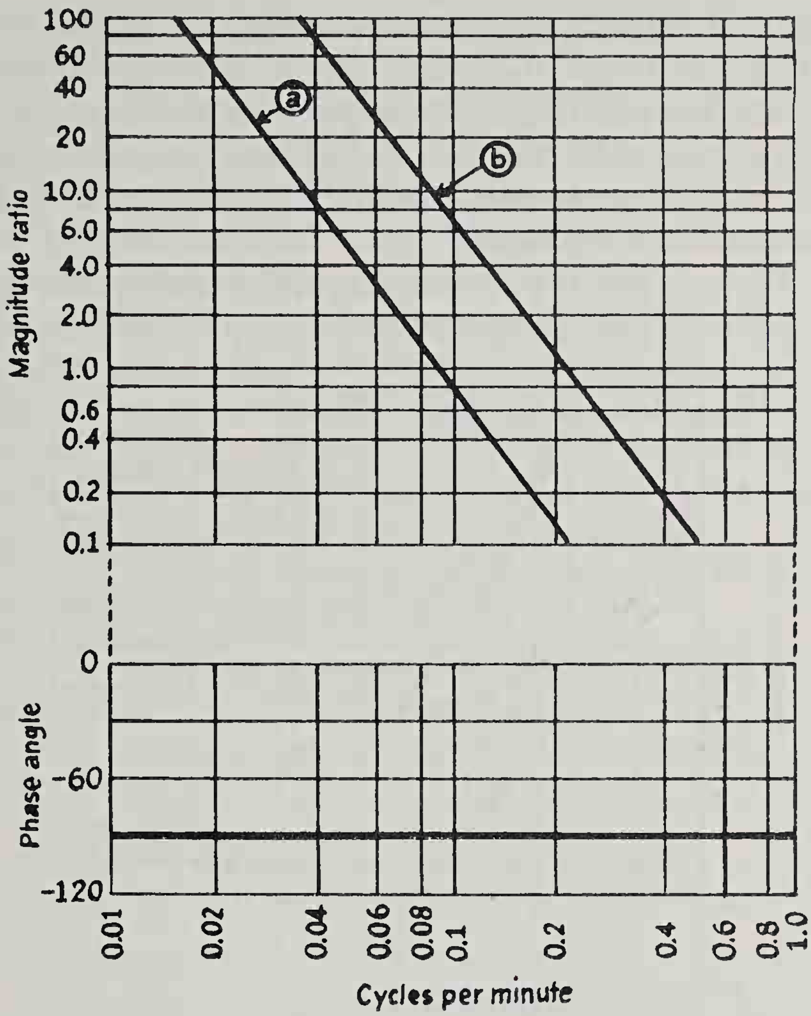


Figure 5–11. Frequency response of proportional-speed floating action.

input signal is slower than the floating speed of the controller, then the amplitude of the output becomes larger than that of the input. Hence the slower the frequency of the signal, the greater the magnitude ratio.

The output of a proportional-speed floating controller reverses direction when the input cycle passes through zero. This is equivalent to a phase lag of 90 degrees and accounts for the steady phase angle in Figure 5–11.

Two graphs, (a) and (b), are shown in the illustration for the magnitude ratio. They correspond to different floating rates of the controller. The magnitude ratio at any given frequency is directly proportional to the floating rate. For example, graph (b) corresponds to a floating rate ten times that of graph (a). The phase angle remains unaffected.

A process without self-regulation behaves essentially like a proportional-speed floating controller. For example, water in a tank with a constant flow leaving the tank will continue changing in one direction for a given flow into the tank. The phase lag would be likewise 90 degrees.

If a process without self-regulation is combined with a proportional-speed floating controller the phase lag of the combination will be 90 + 90 = 180 degrees, and at some point of the frequency response curve the magnitude ratio will be 1.0. Hence, this is an unstable combination. The proportional-speed floating controller cannot be used for processes without self-regulation.

Figure 5–12 illustrates the action of a proportional-speed floating controller in a control loop. In this case, the loop is considered to consist

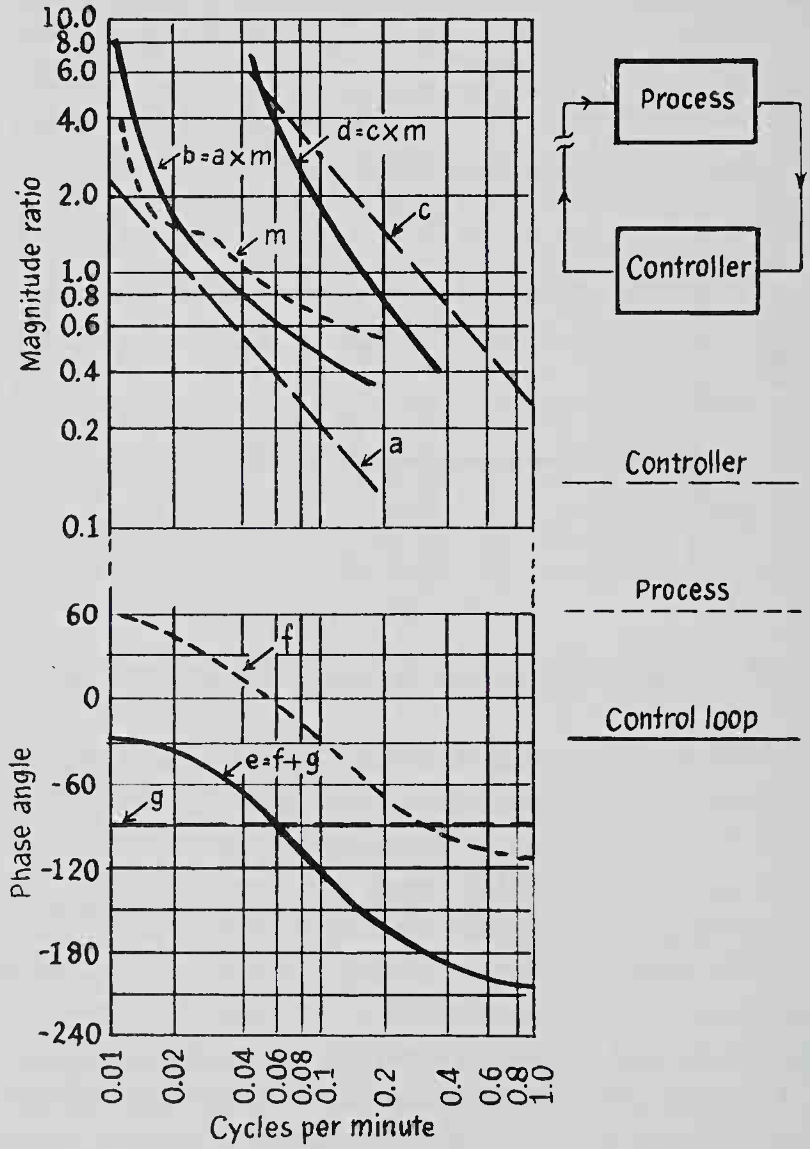


Figure 5–12. Proportional-speed floating action in control loop.

only of controller and process. Graphs (a) and (c) are controller responses; graph (m) is the process response; graphs (b) and (d) are the control loop responses of process and controller combined. Initially, the floating rate of the controller is set for condition (a) and the response of

the control loop (b) results from combining it with the frequency response of the process.

The phase angle graph (e) is the sum of the phase angles of process and controller. At approximately 0.17 cycle/min., the phase angle is −150 degrees. Comparison with the magnitude ratio curve shows that it is possible to increase the floating rate of the controller considerably and still remain within the conditions of stability.

Graph (c) corresponds to a floating rate about ten times that of (a). The new control loop response is illustrated by graph (d). This response has a magnitude ratio of 1.0 at 0.17 cycle/min., i.e., at the frequency at which the phase angle is −150 degrees. At −180 degrees the magnitude ratio is 0.5. Stability is thus still assured and the speed of response of the controller has improved considerably.

Proportional-Position Controller

The frequency response curve of a proportional-position controller would be a straight line, both for the magnitude ratio and the phase angle up to relatively high frequencies. This is illustrated in Figure 5–13.

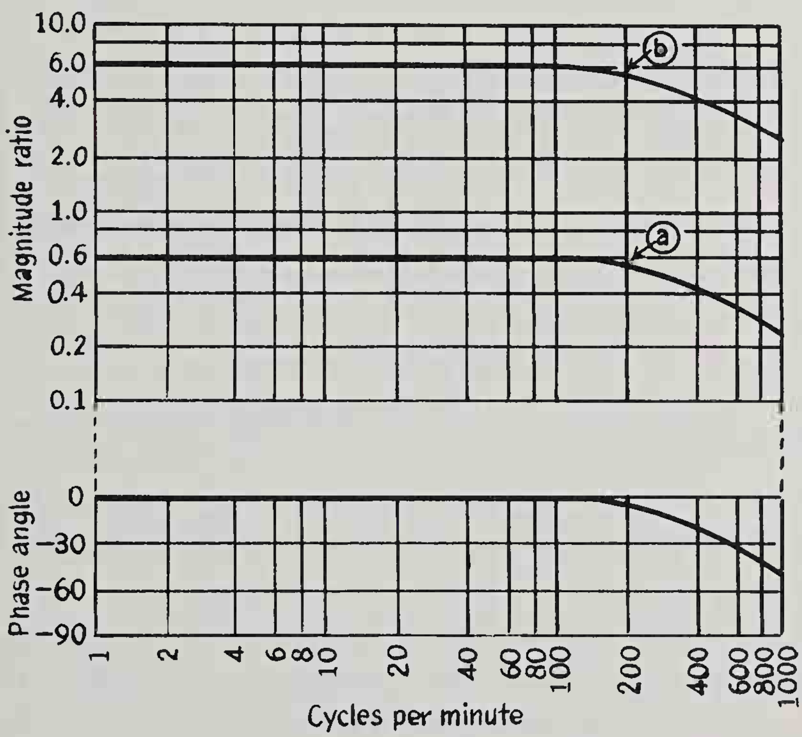


Figure 5–13. Frequency response of proportional-position action.

The principle of proportional-position action implies that a change of the input signal is reflected in an immediate output signal of proportional magnitude. However, when the input signal changes very rapidly, the output signal can no longer follow, and attenuation of the signal as

well as delay occurs, which produces the decrease of magnitude ratio and the increase of phase lag.

Graph (b) corresponds to a magnitude ratio ten times that of (a). This is produced by a change in proportional band. Narrowing the proportional band of the controller to one tenth increases the magnitude ratio by ten. The phase lag is not altered.

Figure 5–14 is the response of a control loop containing a proportional-position controller and a process with dead time only. Graph (a)

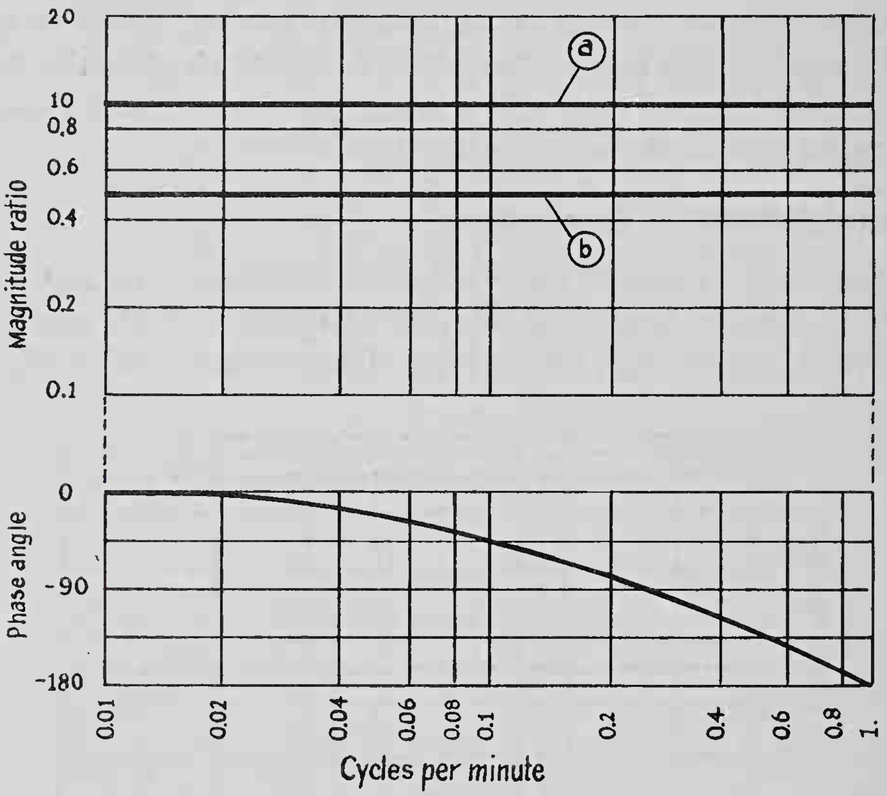


Figure 5–14. Proportional-position action with dead time.

represents an unstable condition since at a 180 degree phase lag the magnitude ratio is not 0.5 or less. If the dead time cannot be eliminated, it is necessary to increase the proportional band.

Suppose the initial magnitude ratio is 1.0 for the process and the controller, respectively. Hence the combined ratio is also 1.0. Doubling the proportional band results in a magnitude ratio of 0.5 for both the controller and the control loop. This corresponds to graph (b) in Figure 5–14 and assures stability.

The widening of the proportional band results in slower control response. It is the compromise that any control system requires: greater stability is equivalent to slower corrective action.

Steady-state Conditions. The condition which corresponds to a frequency signal that approaches zero cycles is called the *steady state.* It is equivalent to a control system that reestablishes balanced conditions after a disturbance.

The magnitude ratio of a proportional-position controller at sufficiently low frequencies is 1.0 with 0 phase angle. This implies that for zero input the output will also be zero.

Offset Characteristics. Offset is an inherent steady-state characteristic of the proportional–position controller. It cannot be overcome because the output-input ratio of the proportional-position controller does not increase at steady state and hence no additional impetus can be expected. This is different from the proportional-speed floating controller which shows an increase of magnitude ratio as the system approaches steady-state condition. Comparison of Figures 5–11 and 5–13 illustrates the difference between the two control actions: one graph is of negative slope, the other is of zero slope.

Proportional-position action, as compared with proportional-speed floating action, offers the advantage of providing an inherently stable component and permitting operation at larger magnitude ratios. This increases the speed of response.

A combination of proportional-speed floating and proportional-position actions, i.e., a proportional plus reset controller, becomes therefore desirable, since it gives fast action and eliminates offset.

Proportional plus Reset Controller

Figure 5–15 shows the response of a proportional plus reset controller. The reset rate of the controller is expressed by the so-called "break frequency." This is illustrated by the dashed lines in Figure 5–15. Both the slanted line, which corresponds to the reset action, and the horizontal line, which corresponds to the proportional-position action, are prolonged as straight lines. The frequency at which these two lines intersect is the break frequency. The phase lag is 45 degrees at the break frequency.

The reset rate is obtained by multiplying the break frequency by 2π or 6.28. This means that

$$I = 6.28f$$

where I is the reset rate in repeats/min. and f is the frequency in cycles/min. Thus, in Figure 5–15, graph (a), the break frequency is 0.1, and hence the reset rate is $6.28 \times 0.1 = 0.628$ repeats/min.

When the reset rate is changed the angle which the slanted line forms

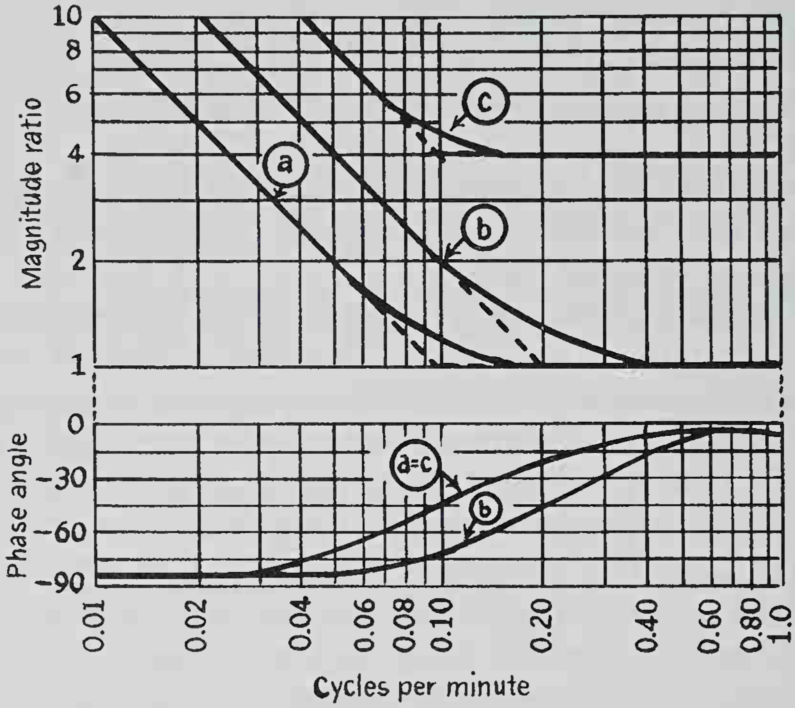


Figure 5–15. Frequency response of proportional plus reset action.

with a horizontal line remains unchanged. The slope of this line always expresses a decrease of the magnitude ratio by roughly 9/10 for an increase of frequency by ten times.

The magnitude ratio graph (a) can be shifted in either the horizontal or the vertical plane. Shifting it in the horizontal plane means changing the reset rate. For example, graph (b) results from an increase of reset rate to 1.25 repeats/min. Similarly, by narrowing the proportional band to ¼, i.e., increasing the magnitude ratio four times, a horizontal shift of graph (a) is obtained and graph (c) results.

The phase angle is unaffected by changes of the proportional band, but it shifts when the reset rate is adjusted. This has some effect on the final adjustment of the controller. It is, however, of little practical consequence, as will be illustrated in the following example, and can usually be ignored.

Adjustment of Proportional plus Reset Controller. In the previous chapter it was described how the reset action of a proportional plus reset controller can be adjusted by using the relation:

$$I = \frac{1.2}{T}$$

where I is the reset rate, and T is the time between two successive maxima of oscillations at the critical setting.

The critical setting which was defined as the widest proportional band

at which continuous oscillations occur, corresponds to a magnitude ratio of 1.0 at a phase lag of 180 degrees. The time T is the period, i.e., the inverse of the frequency, at this condition.

On this basis, it is possible to determine the controller adjustments from the frequency response data of a control loop. These data should be taken with the reset action omitted.

Suppose the results are those of Figure 5–16 where the magnitude ratio is 2.0 for frequencies up to about 0.04 cycles/min. In order to

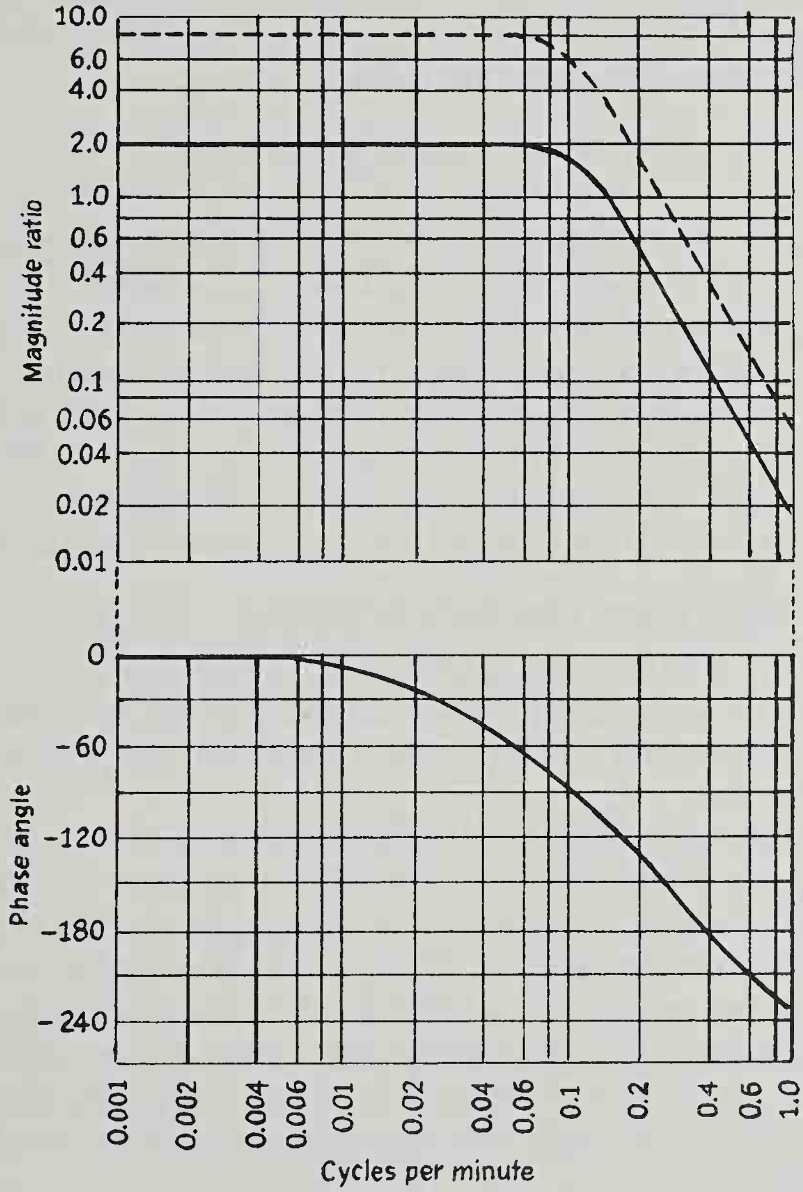


Figure 5–16. Frequency response of control loop without reset.

determine the proportional band, the frequency at a phase angle of −150 degrees is determined. In this case, it is about 0.27 cycles per minute. In accordance with the rule 1 of the stability conditions, the

graph of the magnitude ratio is shifted in a vertical plane until a ratio of 1.0 coincides with this frequency. The dashed line corresponds to this condition.

It will be found that the new amplitude ratio is four times the old one at any frequency. This means that the controller sensitivity can be increased by a factor of four, which is equivalent to reducing the proportional band to ¼ its previous value.

To ascertain the reset rate, it is necessary to determine the frequency at a 180 degree phase lag, which in the present case is 0.4 cycles/min. Since the period is the inverse of the frequency this corresponds to a period of 2.5 min. The reset rate is then

$$I = \frac{1.2}{2.5} = 0.48 \text{ repeats/min.}$$

Figure 5–17 illustrates the change of the control loop response due to the addition of reset action. The condition of the dashed line of Figure 5–16 is combined with the added reset. The phase angle is represented in three graphs. One, (a), is the response without reset action; (b) is the phase angle due to proportional plus reset action; and (c) is the resulting combination. Although the (c) graph differs considerably from the original (a) graph, both show the phase lag of 150 degrees at practically the same frequency. Hence the stability condition remains unaltered.

Proportional plus Reset plus Rate Controller

The addition of rate action provides corrective action in proportion to the speed with which a controlled variable deviates from its set point. It is used to compensate for large time constants in the process or in the measuring element.

Since the effect of rate action is proportional to the speed of deviation, it must also increase its effect with the frequency of a test signal. Figure 5–18 shows the frequency response of a proportional plus reset plus rate controller. The increase of the magnitude ratio after about 0.3 cycles/min. is due to rate action. The graph has two break frequencies: one at 0.08 cycles/min., the other at 0.4 cycles. From the first break frequency the reset rate can be determined as previously described. In this case it is 0.5 repeats/min. The rate time can be determined by using the equality:

$$D = \frac{1}{2\pi f} \quad \text{or} \quad D = \frac{0.16}{f}$$

where D is the rate time in minutes and f is the frequency in cycles/min. Thus in the present case

$$D = \frac{0.16}{0.4} = 0.4 \text{ min.}$$

The slope which the slanted line forms with the horizontal is the same whatever the rate time is. Its magnitude is also the same as the slope of the reset action, only it is positive instead of negative. Increasing, for example, the rate time to 0.8 min., shifts the right-hand slanted portion of the graph to the left until a break frequency of 0.2 cycles/min. is obtained.

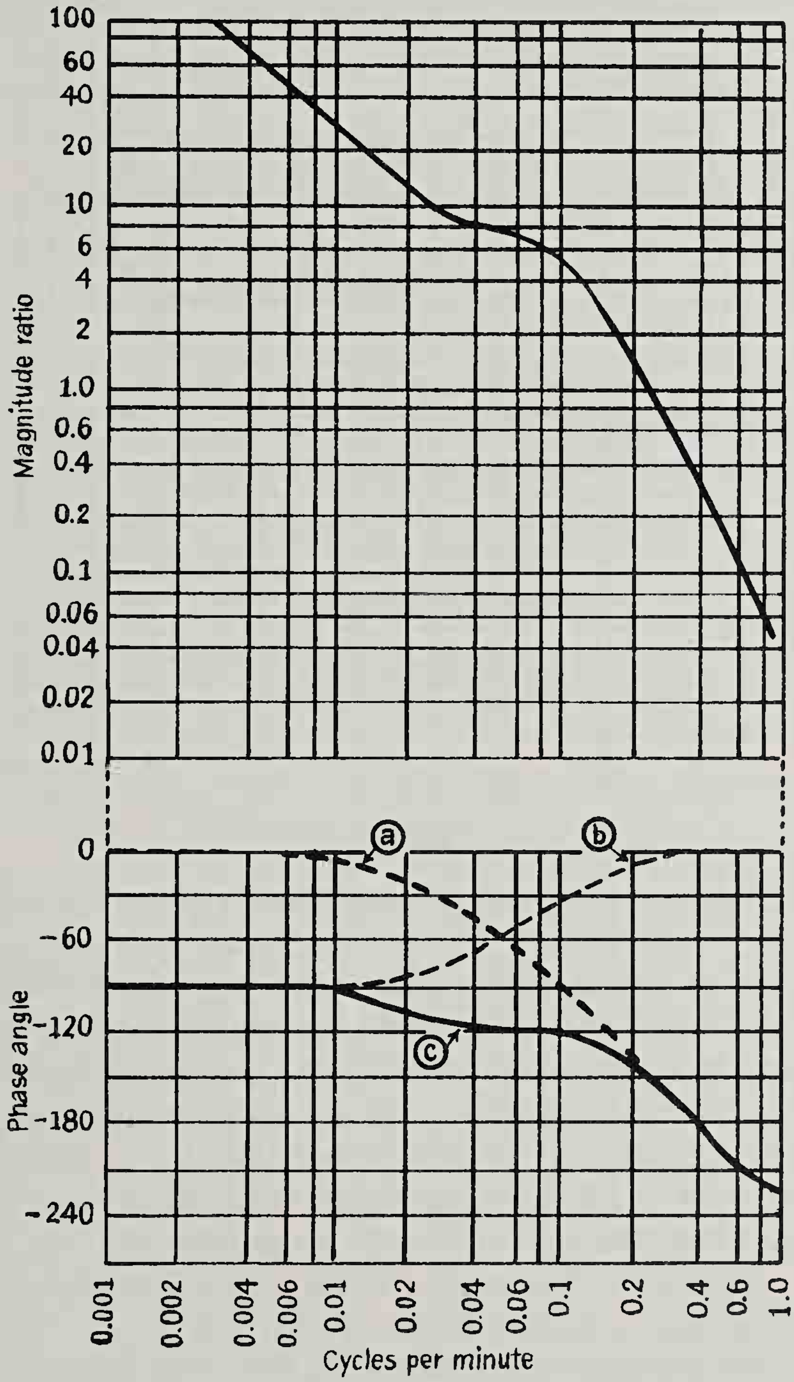


Figure 5–17. Frequency response of control loop with reset.

By changing the proportional band, the entire graph is shifted in a vertical plane. For example, by cutting the proportional band in half, graph (b) is shifted to become graph (a).

The phase angle becomes modified by the added rate action, and the output signal leads the input signal at frequencies that are under the influence of rate action. In Figure 5–18, this starts at about 0.2 cycles/min.

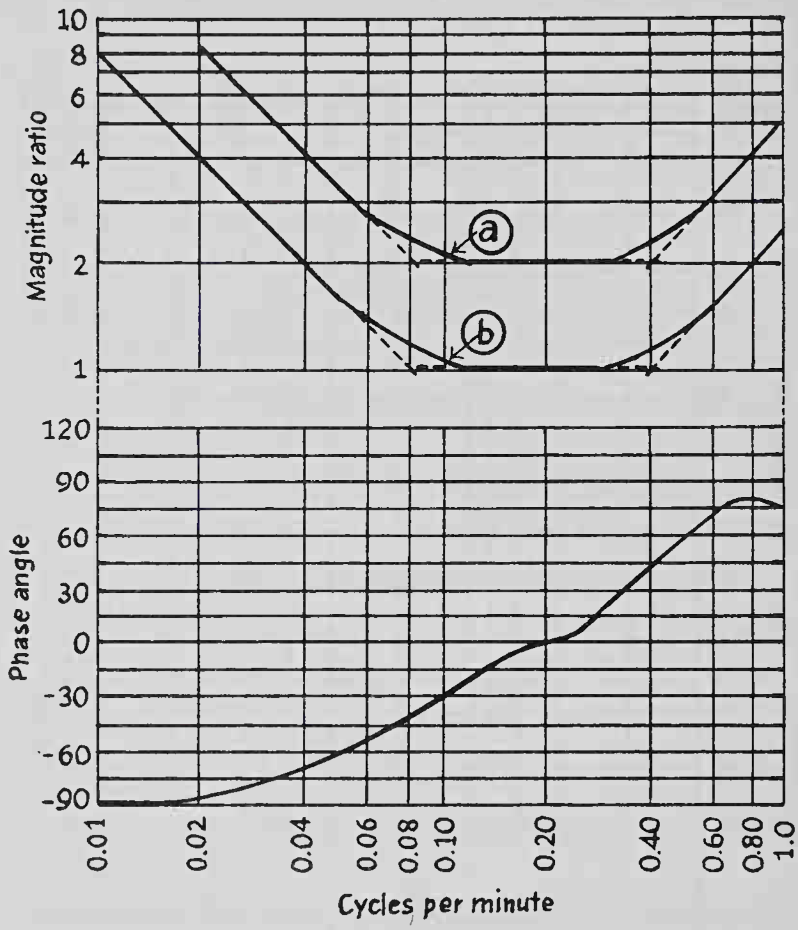


Figure 5–18. Frequency response of proportional plus reset plus rate action.

Adjustment of Proportional plus Reset plus Rate Controller. In the previous chapter which discussed the adjustments on the basis of step function response, it was pointed out that the proportional band should be twice the critical setting for proportional-position controllers with or without reset. It was mentioned that the proportional band could be narrowed to 1.7 times instead of twice the critical setting when rate action is added. This means that the magnitude ratio can be increased because of rate action. The reason is that the effect of the phase shift becomes noticeable, as will be shown further below.

For the same reasons, when adjusting from the frequency response as in Figure 5–19, the magnitude ratio of 1.0 no longer refers to a phase

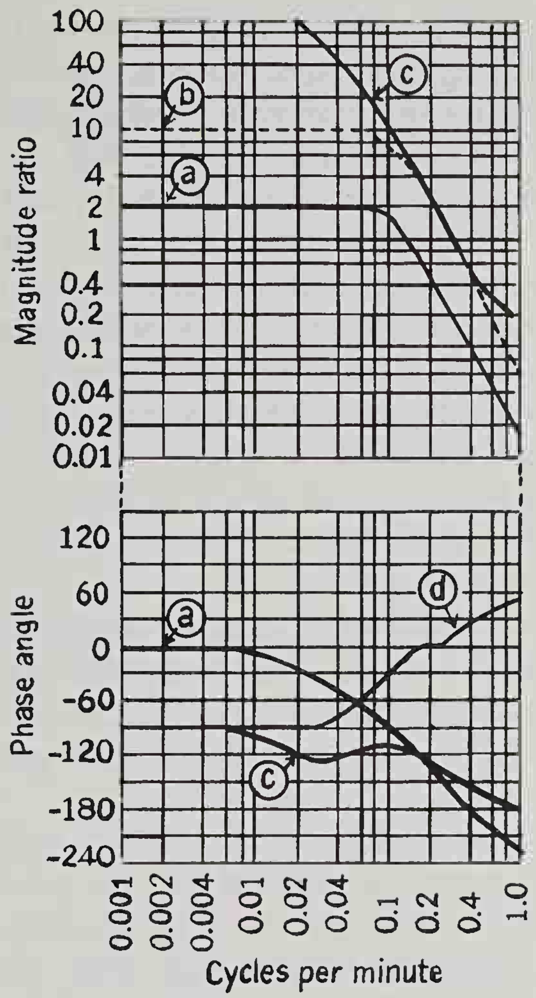


Figure 5–19. Frequency response of control loop without and with rate and reset.

angle of −150 degrees, but to one of −160 degrees. Figure 5–19 shows in graph (a) the same frequency response as in Figure 5–16. Reset and rate actions are cut off. Adjusting the response of Figure 5–19 for minimum proportional band at stable conditions results in the dashed line (b). The magnitude ratio is now 1.0 at −160 degrees and at low frequencies it is 10.0 instead of 2.0, which is 5 times the initial ratio or 0.2 times the initial proportional band.

To determine the reset and rate actions, the corresponding expressions of the previous chapter are used. They were:

$$I = \frac{2}{T}$$

and

$$D = \frac{T}{8}$$

where I is the reset rate in repeats/min., D is the rate time in minutes

and T is the period in minutes referred to the critical setting, i.e., to a -180 degree phase angle. In the present case, the period at -180 degrees is 2.5 min. Hence the reset rate is 0.80 repeats/min. and the rate time is 0.31 min.

Adjusting reset and rate actions accordingly, and then repeating the frequency response test of the control loop, results in graph (c) of Figure 5–19 both for magnitude ratio and phase angle. The phase angle in graph (c) is the result of adding the process phase angles (a) to the controller phase angles (d). It can be seen that at the frequency at which curve (a) showed -160 degrees, the phase angle now shows only -150 degrees. This justifies the initial assumption of 160 degrees in case of rate action instead of the conventional 150 degrees.

Conclusions

The preceding description was illustrated by graphs which were more or less idealized. Actual readings frequently show considerable random variations for which a smooth curve only indicates a trend. In view of these conditions a differentiation between 150 and 160 degrees may become impractical.

The question arises as to the accuracy of this method. This points to the limitation of frequency response data used for the setting of automatic controllers. For a system which is actually installed, the step function response method is simpler and more reliable than the frequency response method. The advantage remaining is that frequency response data allow the theoretical combination of several components without their being installed, and the estimation of what approximate magnitude their settings will be.

6. MECHANICAL COMPONENTS

The measuring means of a controller converts the magnitude of the controlled variable into either a mechanical or an electrical signal. In the majority of cases, the mechanical input signal is a pressure which is applied to a diaphragm, bellows, or Bourdon tube to convert it into a force or position. Occasionally force or position is obtained without conversion, as in the deflection of a bimetallic element or the displacement of a float in a level controller. However, even then, a secondary conversion into a mechanical or electrical signal is usually desirable.

Mechanical systems generally use hydraulic or pneumatic fluids. Figure 6–1 shows the most common components of such control systems. In a pneumatic system, for example, the deflection of the bellows positions a flapper of a flapper-nozzle combination. This results in a change of the pressure of a secondary air supply which is applied to the

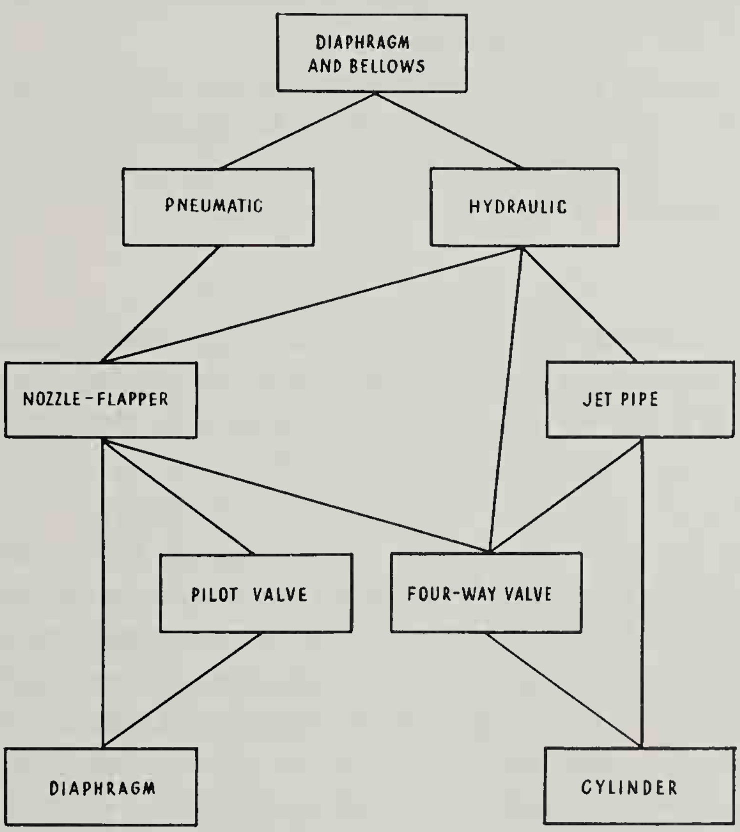


Figure 6–1. Components of pneumatic and hydraulic control systems.

diaphragm of a final control element. In order to amplify the output from the flapper-nozzle, a pilot valve may be connected between it and the final diaphragm. In some cases the pressure which is applied to the bellows as input signal to the controller can be used directly at the diaphragm of the final control element. The resulting simplification is desirable. However, in many cases this primary signal would not be powerful enough. It would furthermore limit the possibilities of control actions such as proportional-position, reset, and rate controls.

In a hydraulic system, the flapper-nozzle may be used in connection with a four-way valve or, if the input force is high enough, the flapper-nozzle stage may be eliminated and the four-way valve be positioned directly. The cylinder's actuating piston, which is linked to the final control element, is positioned through the four-way valve. Another device is the jet pipe which either positions an actuating piston directly or acts through a four-way valve.

Internal feedback arrangements produce numerous interconnections and repetitions of all these components. Thus a flapper-nozzle may be used in the measuring means, again in the controller, and again in the positioning of the final control element.

The description of various components in the succeeding pages follows largely the order outlined in Figure 6–1. An exception is made with one element which is probably the most universally used. This is the *spring* and it will be discussed first.

Springs

One of the most universal elements in mechanical control elements is the spring. The helical spring has a deflection which is proportional to the force applied. Written as an equation, this means that

$$F = kx$$

where F is the force in lb, k is the spring rate in lb/in. deflection, and x is the deflection in inches. This equation is known as Hooke's law. It is valid within the elastic limits of the spring. The practical behavior of the spring does not necessarily follow completely the idealized Hooke's law. Deviations from linearity of as much as 5 per cent of the total deflection must be expected.

The spring rate is given by

$$k = \frac{Gd^4}{8D^3N}$$

where G is the torsional modulus of rigidity, which for steel wire is approximately 11,500,000; d is the diameter of the spring wire in inches,

as shown in Figure 6–2; D is the coil diameter in inches, as shown; and N is the number of active coil. The equation shows that minute changes of coil diameter and even smaller changes in spring wire diameter will

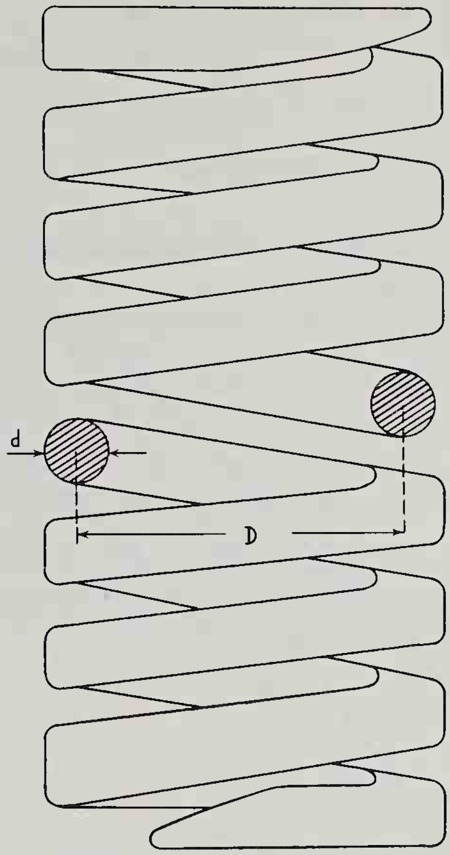


Figure 6–2. Helical spring.

produce large changes in the spring rate. This is the main reason for the difficulty of controlling spring rates in production. The usual method is to select from a production run those that come closest to a specified spring rate, and even then deviations of 5 per cent of the spring rate are generally accepted.

The above equation also shows that the spring rate is inversely proportional to the number of coils. It is therefore possible to increase spring rates by cutting out coils. For example, decreasing the number of coils by 50 per cent increases the spring rate twofold.

Helical springs are either compression springs or tension springs, as shown in Figure 6–3. The above mentioned equations are valid for either form. The deflection x refers to compression as well as tension.

Figure 6–4 illustrates the difference between spring rate and spring force. Graph (1) illustrates a spring rate of 100 lb/in. If this corresponds to a spring with 7 coils, then the same spring with 10 coils would have a spring rate of 70 lb/in., as illustrated in graph (2). Using, however,

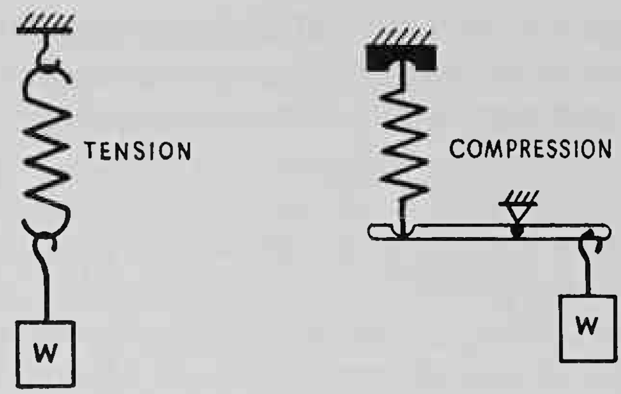


Figure 6–3. Tension and compression springs.

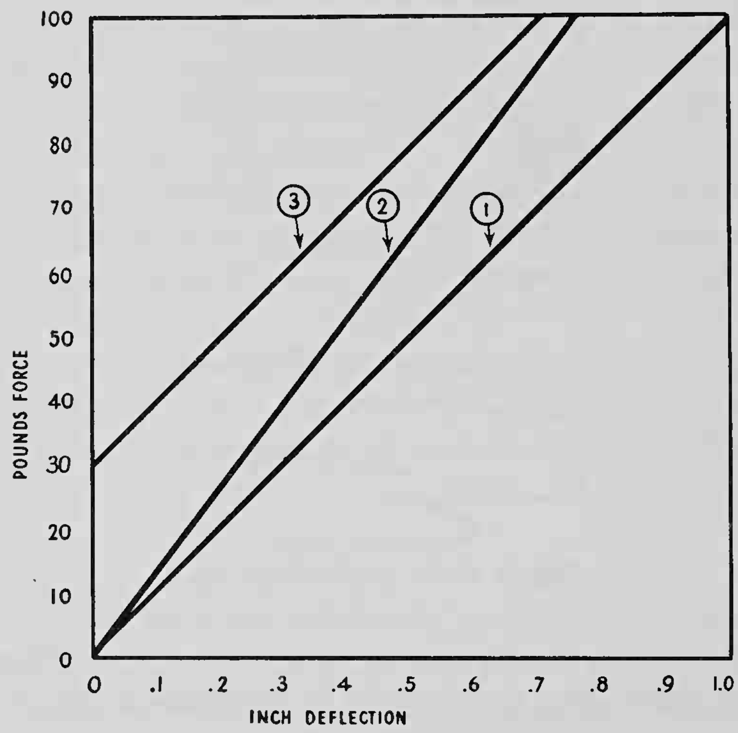


Figure 6–4. Spring rate and spring force.

the 7-coil spring and precompressing the spring by 0.3 in. increases the spring force correspondingly, as shown in graph (3).

Springs may be combined either in series or in parallel arrangement. The basic combinations are those seen in Figure 6–5. Case *a* is a series connection of two coils with spring rates k_1 and k_2. The combined spring rate k is smaller than that of either k_1 or k_2. It can be calculated by the equation

$$k = \frac{1}{\frac{1}{k_1} + \frac{1}{k_2}}$$

Cases *b* and *c* are parallel connections. Since each coil acts on the weight

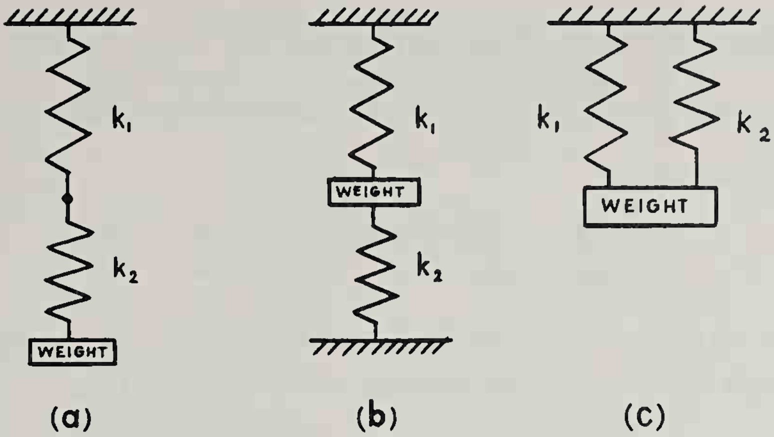


Figure 6–5. Spring combinations.

directly, the total spring rate is the sum of the individual spring rates, or

$$k = k_1 + k_2.$$

Frequently, the springs are used in connection with levers, as shown in Figure 6–6. The force which is exerted by the weight on the spring in case

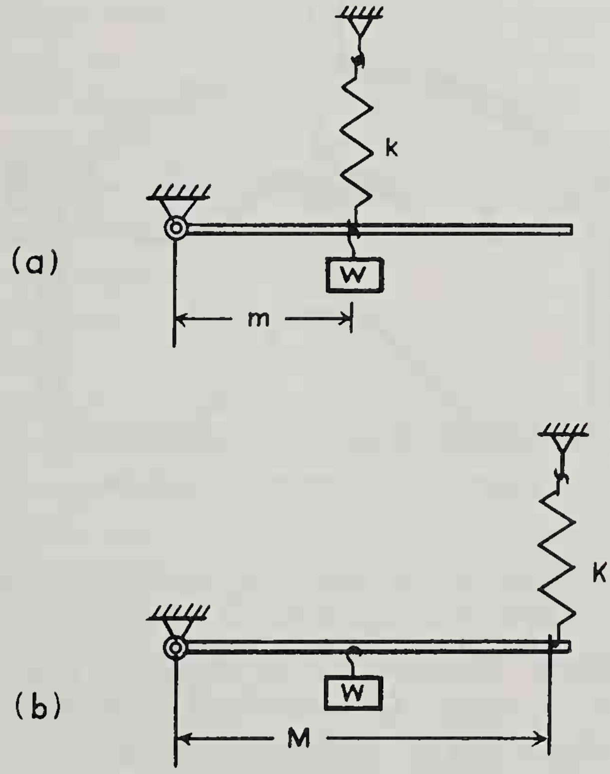


Figure 6–6. Spring on lever.

a is *W*, which is balanced by the spring with a force equal to kx. Increasing the lever of the spring force as in case *b* increases the acting force

by $\frac{M}{m}$, and the deflection of the spring for the same motion of the lever also increases by $\frac{M}{m}$. Hence the weight now becomes balanced by a spring force which is equal to

$$\frac{M}{m} \cdot k \cdot \frac{M}{m} \cdot x, \quad \text{or} \quad F = \frac{M^2}{m^2} kx$$

which shows that the effect of the lever is to increase the spring rate by a factor which is equal to the square of the lever ratio.

Diaphragms

Pressure applied to an area is force, i.e., $P \cdot A = F$. This means that once the pressure is given, the force can be derived by knowing the effective area to which the pressure is applied. If a diaphragm is used, the effective area is less than the actual area of the diaphragm. The pressure that acts on the diaphragm, as shown in Figure 6–7, produces a force

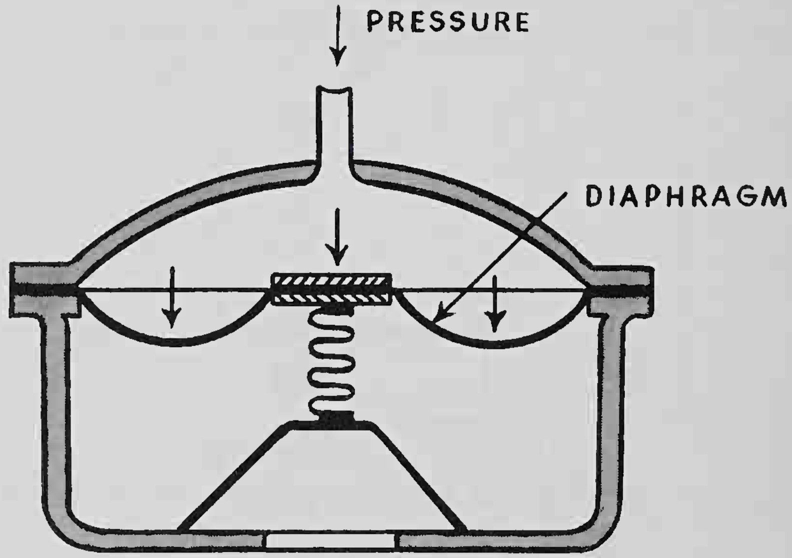


Figure 6–7. Diaphragm system.

that is absorbed in part by the clamping of the periphery of the diaphragm. The remaining force is available to compress the spring.

Slack Diaphragms. Diaphragms of sheep or goatskin leather are probably the most sensitive. However, synthetic materials, generally with a fabric layer, have largely superseded the leather diaphragms. The convolution is either molded into the material, or sufficient slackness is provided so that the convolution is formed by the pressure loading.

Considering the lowest part of the convolution of the diaphragm, the effective diameter of the diaphragm changes as illustrated in Figure 6–8.

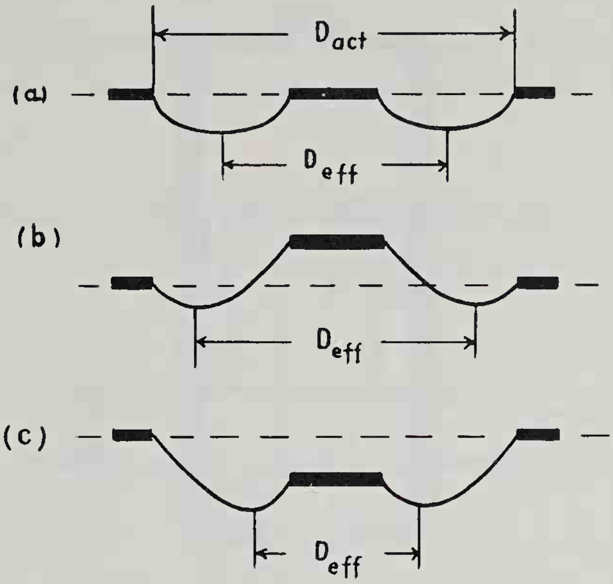


Figure 6–8. Various deflections of diaphragm.

The mid position is illustrated in Figure 6–8a, while in 6–8b the pressure has decreased. This moves the diaphragm upward due to the force of a spring which is not shown, and changes the shape of the convolution. As a consequence of the new shape of the convolution, the effective diameter increases. Conversely, in Figure 6–8c the pressure has increased, the diaphragm has moved downward and the convolution is now of such shape that the effective diameter has decreased.

Area changes with the square of the diameter, hence a relatively small change in effective diameter will have a considerably more pronounced effect on the area and hence on the force that is produced by the pressure applied to the diaphragm.

The change in effective diameter and hence of the effective area has the result that the force is no longer proportional to the pressure. A nonlinear element is the result. In order to minimize this nonlinear effect, the stroke should be reduced to a minimum. This is not always possible, particularly in spring-opposed diaphragm valves where strokes are frequently 2 or more inches. In such cases, the convolutions are made relatively deep to reduce the shifting of the effective diameter. This has a side effect insofar as it produces a slightly different shape of convolution depending on whether the valve moves upward or downward. The consequence is the equivalent of backlash.

"Belloframs." In order to increase the linear effect of diaphragms, the Bellofram Corporation developed a particular shape of diaphragm, called the "Bellofram," in which the deep convolution of the molded synthetic diaphragm rolls between two surfaces, as illustrated in Figure 6–9. The effective area is maintained constant for all practical purposes.

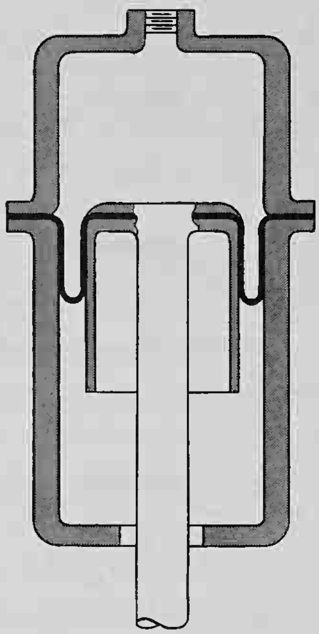


Figure 6–9. "Bellofram."

Metal Diaphragms. Slack diaphragms have a pressure rating that generally limits them to pressures of 20 psi or less. Metal diaphragms are less limited in this respect. Their structural rigidity is of advantage in designing components. Generally, they can be used without counteracting springs.

Metallic diaphragms may cause nonlinearities because of their rapid change of effective area with deflection. This can be made negligible by reducing the deflection. Flat plate diaphragms must not be deflected by more than about ⅓ of their thickness. This means a motion in the order of a thousandth of an inch, which is quite acceptable in many components. In order to improve the linearity over somewhat longer strokes, corrugated metal diaphragms are frequently used. The corrugations are generally concentric.

Bellows

Further increase of deflection, relative rigidity, and comparatively high pressure rating are characteristic of bellows. The amount of deflection per unit change of pressure is a function of the number of corrugations, the spring rate, and the area of the bellows. Thus a typical bellows may have a spring rate of 600 lb/in. corrugation. Providing ten convolutions, a spring rate of 60 lb/in. results. Let the area of the bellows be 0.3 sq in.; with a pressure change of 20 psi, the bellows will then deflect 0.1 in.

Hysteresis and nonlinearities of several per cent of maximum stroke may have to be expected with bellows. To improve performance, a

helical spring is used parallel with the bellows. The spring rate of the helical spring should be ten times that of the bellows.

Flapper-Nozzle

The flapper-nozzle as illustrated in Figure 6–10 is an adjustable area in a flow passage. In this respect it behaves like a valve. The advantage

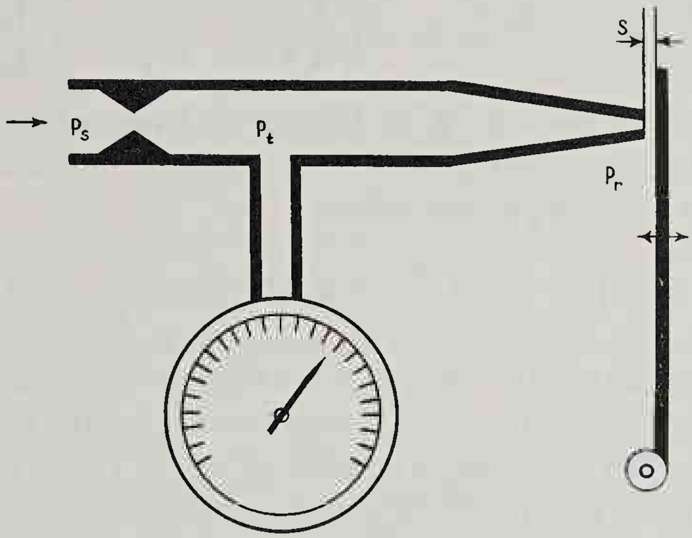


Figure 6–10. Flapper-nozzle.

over a valve is the simplicity of construction and the smallness of force required to position it. The adjustable flow restriction of the flapper-nozzle is used in series with a fixed restriction. Supply air enters the fixed restriction at an absolute pressure p_s. It drops from p_s to p_t in passing through the fixed restriction, and from p_t to p_r in passing through the nozzle. A restriction with a passage of negligible length is considered an orifice. This applies to both fixed and adjustable restrictions. The flow which comes through the fixed restriction leaves through the nozzle and the rate of air flow Q can be approximately expressed by

$$Q = kA_1 \sqrt{(p_s - p_t)p_t} = kA_2 \sqrt{(p_t - p_r)p_r}$$

All pressures are absolute pressures, A_1 and A_2 are the areas of fixed restriction and nozzle, respectively, and k is a constant which depends on the flow pattern and the conditions of the fluid. The above equation can be written

$$A_1 \sqrt{(p_s - p_t)p_t} = A_2 \sqrt{(p_t - p_r)p_r} \tag{6–1}$$

After squaring and rearranging, the following equation results:

$$p_t^2 + \left(\frac{A_2^2}{A_1^2} p_r - p_s\right)p_t - \frac{A_2^2}{A_1^2} p_r^2 = 0$$

Solving this equation for p_t gives

$$p_t = \frac{1}{2}\left[p_s - \frac{A_2^2}{A_1^2}p_r \pm \sqrt{\left(\frac{A_2^2}{A_1^2}p_r - p_s\right)^2 + 4\frac{A_2^2}{A_1^2}p_r^2}\right]$$

Only one solution, however, gives positive results and these are the only ones of practical significance. The equation, therefore, becomes

$$p = \frac{1}{2}\left[p_s - \frac{A_2^2}{A_1^2}p_r + \sqrt{\left(\frac{A_2^2}{A_1^2}p_r - p_s\right)^2 + 4\frac{A_2^2}{A_1^2}p_r^2}\right]$$

Suppose the supply pressure p_s is 30 psia and the reference pressure is atmospheric, i.e., approximately 15 psia. Inserting these values in the above equation and simplifying gives

$$p_t = \left(15 - 7.5\frac{A_2^2}{A_1^2}\right) + 15\sqrt{0.25\frac{A_2^4}{A_1^4} + 1} \qquad (6\text{–}2)$$

The area of A_2 is changed by means of the flapper, hence the ratio between A_2 and A_1 can assume any value within certain limits. Suppose that A_2 varies from $0.2A_1$ to $2A_1$, then a number of A_2/A_1 ratios can be assumed within these limits, and the corresponding magnitudes of p_t can be calculated. The nozzle back pressure expressed in gauge pressure is approximately equal to $(p_t - 15)$.

The following is a table of computed values for the above conditions:

A_2/A_1	p_t	Nozzle Back Pressure in psi	A_2/A_1	p_t	Nozzle Back Pressure in psi
0.2	29.3	14.3	1.2	22.7	7.7
0.4	28.8	13.8	1.4	21.4	6.4
0.6	27.5	12.5	1.6	20.3	5.3
0.8	26.0	11.0	1.8	19.4	4.4
1.0	24.3	9.3	2.0	18.5	3.5

These values are plotted as curve in Figure 6–11 and show how the nozzle back pressure varies with the position of the flapper. The curve illustrates the nonlinearity in this relation, but also makes it clear that the linearity can be improved considerably by operating only over a limited range of area ratios, e.g., from $A_2/A_1 = 1.2$ to $A_2/A_1 = 1.4$. This results in correspondingly small pressure changes and is one of the reasons that pilot valves, as described later, are used to multiply the small nozzle back pressure change into a 3 to 15 psi controller output signal range.

Nozzle diameters are in the order of magnitude of 1/64 in. This cor-

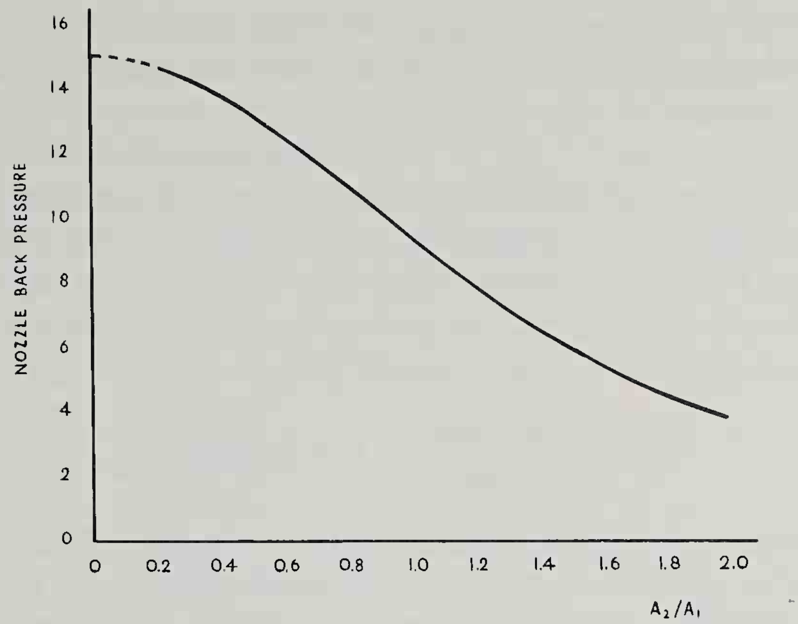


Figure 6–11. Change of nozzle back pressure with flapper position.

responds to an area of the nozzle opening about 0.0002 sq in. Since the force of the air stream impinging on the flapper is approximately equal to $p_2 \cdot A_2$, the maximum force that may result when the flapper completely closes the nozzle is 0.003 lb. Actually, even this minute force is not obtained, because the nozzle is always partially open. On the other hand the simple fact that a force exists points to certain limitations. For example, if the nozzle opening should be increased to ⅛ in., the force would increase 64 times and would no longer be negligible. Similarly, the air pressures that can be handled cannot be increased indefinitely.

Another reason for operating with small nozzle diameters and low nozzle back pressures is to reduce the air consumption. In general, the air consumption is in the order of 1 to 2 standard cu ft/hr. For this purpose the fixed orifice is generally about 1/128 in. in diameter.

The flow passes through the nozzle, impinges on the flapper and is deflected at right angles through an area which is given by the circumference of the nozzle opening and the distance s between flapper and nozzle (Figure 6–10). Considering a fixed orifice of 1/128 in. in diameter, this corresponds with an area of approximately 0.00005 sq in. For a nozzle 1/64 in. in diameter the circumference is about 0.05 in. For an area ratio of $A_2/A_1 = 2$, the distance s would be given by

$$s = \frac{0.00005}{0.05} \times 2 = 0.002 \text{ in.}$$

This clearance is, in fact, so small that it is not practical to close the nozzle tighter than this. Previous calculations have shown that for a supply

pressure of 15 psi (30 psia), the nozzle back pressure for this area ratio is 3.5 psi, which consequently is usually the maximum back pressure in the flapper-nozzle.

In order to obtain 15 psi as maximum controller output signal, an amplification factor of $15/3.5 = 4.3$ is required in the pilot valve. For a minimum output of 3 psi, a nozzle back pressure of 0.7 psi would then be required. From equation (6–2) it can be computed that this nozzle back pressure is the result of an area ratio of approximately $A_2/A_1 = 3$. This needs a distance between flapper and nozzle of

$$s = \frac{0.00005}{0.05} \times 3 = 0.003 \text{ in.}$$

In other words, the flapper motion of $0.003 - 0.002 = 0.001$ in. regulates the controller output signal over its entire range. Smaller ranges of flapper motion, e.g., 0.0006 in., are by no means uncommon.

The fluid for the flapper-nozzle as described here was air or any other gas. The flapper-nozzle can be and is used also for liquids, specifically, hydraulic oil. The characteristics are similar, even though the flow equations as given above do not apply. Hydraulic operation of flapper-nozzles has, however, its difficulties. High pressures occurring with hydraulics produce forces on the flapper which are no longer negligible. Furthermore, it is more difficult to keep oil clean than air, and the extremely small diameters of the flow resistances clog easily, unless all impurities are continuously removed.

Other Forms of Flapper-Nozzles. Figure 6–12 shows a push-pull arrangement of a flapper-nozzle. When the flapper moves toward one

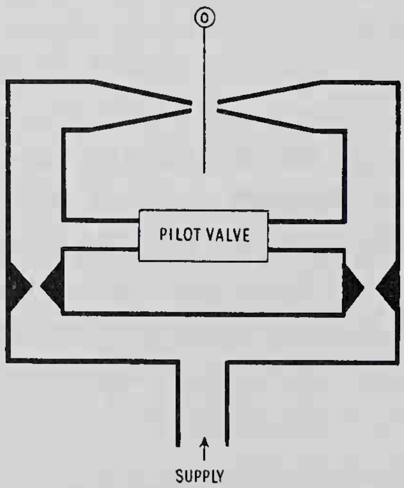


Figure 6–12. Push-pull flapper-nozzle.

nozzle it moves away from the other. The effect is therefore to increase the nozzle back pressure on one and simultaneously to decrease it on

the other. The pilot valve must be designed so that it operates on a pressure differential. The push-pull flapper nozzle is occasionally used in hydraulic circuits where the differential pressure developed is applied across the spool of a four-way valve.

Figure 6–13 shows the free-vane principle used by the Bristol Co. The vane takes the place of the flapper. The main advantage is that flow

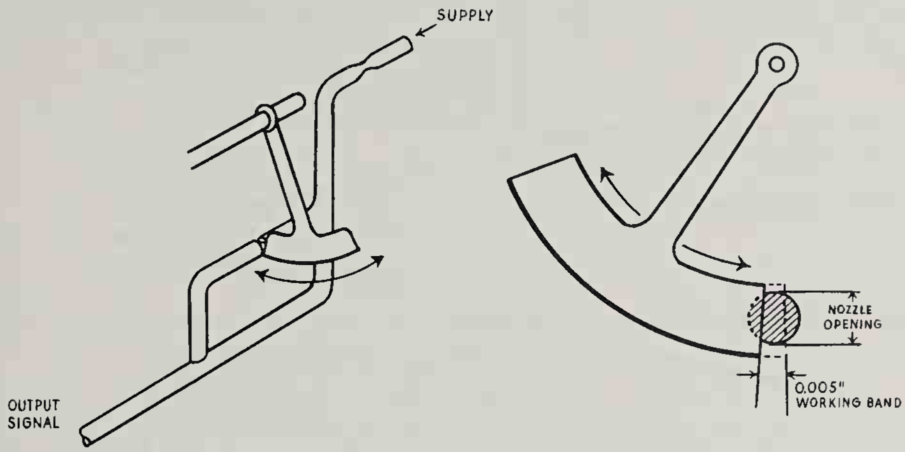


Figure 6–13. Free vane principle. (*The Bristol* Co.)

forces on the vane are at a minimum because the motion of the vane is vertical to the air stream, letting the nozzles and sidewise forces balance each other because of the twin nozzle design. The edge of the vane is between the two nozzles. It operates with a motion of 0.005 in., changing the nozzle back pressure from 4.5 to 5.5 psi.

Other Pneumatic Flow Restrictions. The orifice is only one way of providing a restriction in a pneumatic circuit. Another method consists of inserting a length of tubing of considerably reduced diameter, a so-called capillary. Flow through a capillary follows the equation

$$Q = 10{,}142{,}220\frac{D^4(p_1 - p_2)}{\eta L}$$

where Q is the flow in cu in./min., D is the inside diameter of the capillary in inches, p_1 and p_2 are the pressures in psi before and after the capillary, η is the viscosity of the fluid in centipoise and L is the length of the capillary in inches.

Since in this equation, the diameter appears in the fourth power, it is readily seen that by decreasing the diameter the rate of flow diminishes rapidly. It is also interesting that the flow is inversely proportional to the length of the capillary. This provides the possibility of adjusting the flow rate by lengthening or shortening the capillary.

Another means for providing a flow restriction is the needle valve. Its flow is expressed by

$$Q = 1900\frac{ma^3L^2}{\eta}\frac{(d/L)^2 + (1 + d/L)^2}{1 - 2(d/L)}(p_1 - p_2)$$

where m, L, and d are dimensions shown in Figure 6–14 expressed in inches, the angle a is also shown in Figure 6–14 and is expressed in

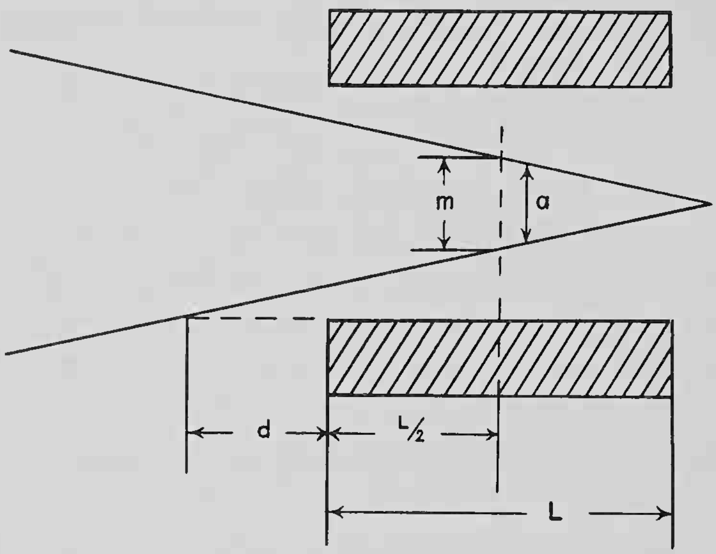


Figure 6–14. Needle valve.

degrees. The viscosity η is in centipoises and the flow Q in cu in./min. Adjustments are usually made with one or two turns of the needle valve handle and cover ranges of 500:1.

Pilot Valves

The pilot valve is a pneumatic amplifier. The flapper-nozzle is designed for small flow capacities, but for fast action of final control elements, larger capacities are needed. Furthermore, the limited pressure range of the flapper-nozzle also requires amplification.

Two basic types of pilot valves are used: the continuous-bleed and the non-bleed type. The continuous-bleed type, illustrated in Figure 6–15, has the advantage of simplicity of construction. The nozzle back pressure is admitted to a pressure-sensitive element; in this case, a diaphragm. A spherical valve plug is connected to the diaphragm by means of a valve stem. In the illustrated position, the exit, which connects to the atmosphere, is closed by the plug and the controller output signal consists of the undiminished air supply pressure. In this position, the nozzle, to which this pilot valve is connected, is obviously wide open since the diaphragm in the pilot valve is not deflected at all. Should the

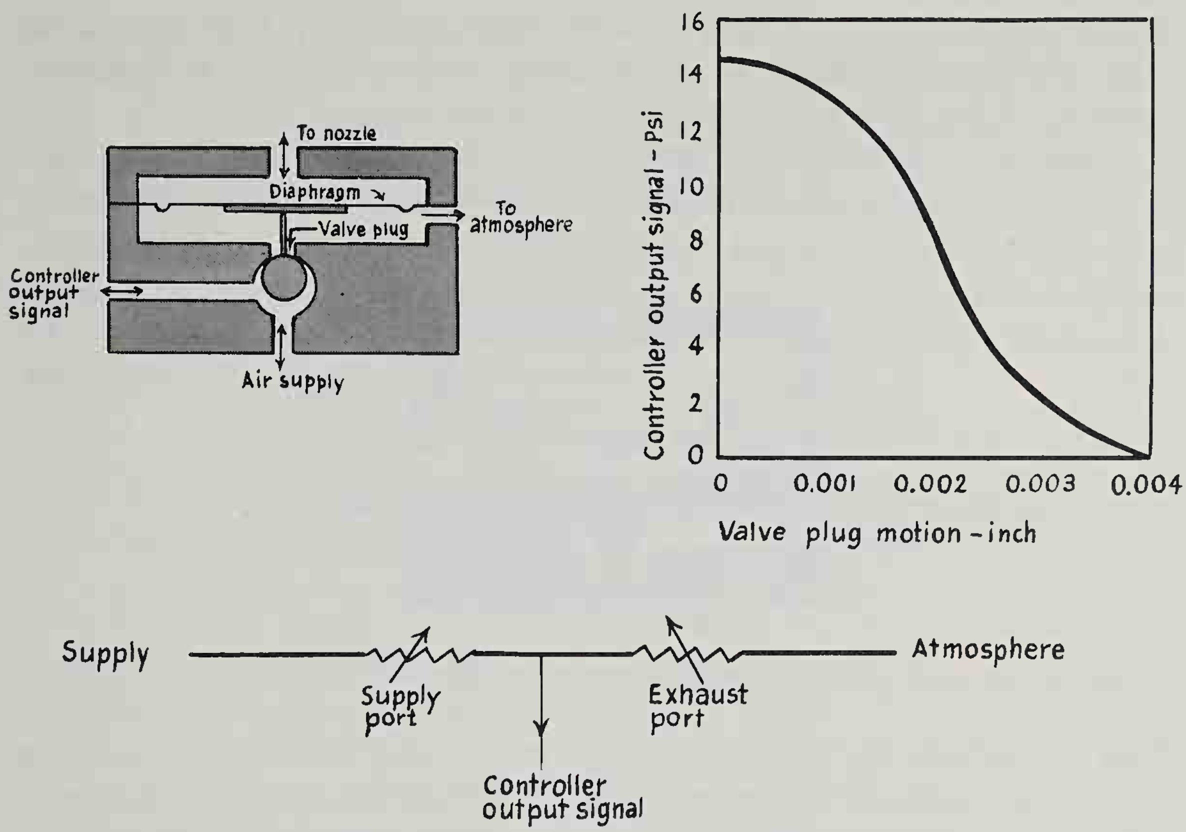


Figure 6–15. Continuous-bleed pilot valve.

nozzle back pressure increase as the flapper moves closer toward the nozzle, the diaphragm would deflect and the valve plug would, at least partially, open the port to atmosphere and close by a corresponding amount the port that admits air from the supply line. The maximum deflection of the pilot valve diaphragm would block off the air supply and release all excess air in the controller output signal line to reduce the signal to atmospheric pressure. All intermediate positions of the pilot valve diaphragm produce intermediate signal pressures in the output.

As shown in the diagram of Figure 6–15, the supply port and the exhaust port are two resistances through which the air may flow. The pressure between the two resistances is the controller output pressure. Supply port resistance and exhaust port resistance may be visualized as being coupled together so that when one resistance increases, the other diminishes automatically, and vice versa.

The nonlinearity between valve plug motion and controller output signal is considerable as shown in Figure 6–15. Hence, in general it is used in this form only for on-off control, when the valve plug assumes either a fully open or a fully closed position. Satisfactory linearity, however, can be obtained by feedback methods. The general procedure is to convert the controller output signal into a position by deflecting a

bellows in proportion to the pressure. This position is then compared with the flapper position as will be described in Chapter 8. The non-linearity of the pilot valve has no effect in this case.

Figure 6–16 illustrates the action of the non-bleed type. Suppose the nozzle back pressure increases due to the flapper approaching the nozzle.

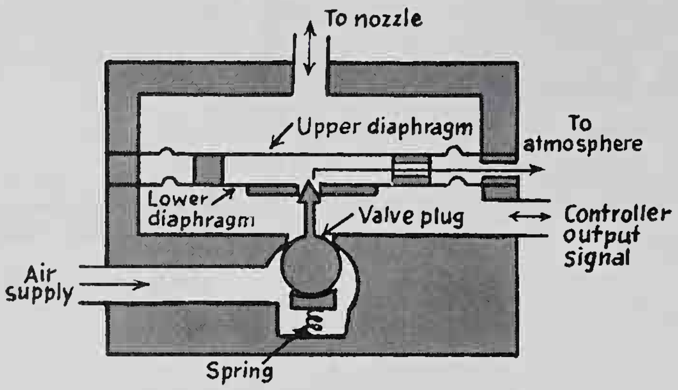


Figure 6–16. Non-bleed pilot valve.

This will deflect the upper diaphragm of the pilot valve and with it the lower diaphragm. It is assumed that no amplification of the change in nozzle back pressure takes place in the pilot valve, although all that would be required for this purpose is a lower diaphragm of smaller area than the upper one.

As the diaphragm assembly moves downward, the valve plug is also pushed down. This action results in an opening of the valve port and admission of supply air to the signal line. The signal pressure which is now rising pushes against the lower diaphragm. Once the signal pressure has increased sufficiently to equal the nozzle back pressure applied to the upper diaphragm, the forces across the diaphragm are in balance, the valve is closed again, and all conditions are the same as shown in Figure 6–16; but the signal pressure of the controller output is now at a higher value. Nonlinearity in the diaphragm response and in the characteristic of the pilot valve does not influence the action.

When the nozzle back pressure decreases, the diaphragm deflects upward. This opens the outlet port to atmosphere and the signal line releases air through this port until the force balance across the diaphragm is reestablished, and the outlet port to atmosphere is again closed.

The non-bleed pilot valve requires a precompressed spring to press the valve plug against the ports as illustrated in Figure 6–16. The force exerted by the nozzle back pressure to deflect the diaphragm downward must be equal to the signal pressure times the effective area of the diaphragm plus the force that results from the precompression of the spring. If, however, the nozzle back pressure decreases and the diaphragm

moves upward, the acting force is only that of the controller output signal applied to the lower diaphragm. The spring force does not enter.

In addition to the spring force there are pressures active across the supply and the exhaust plugs. Calling the supply pressure p_s, the controller output signal pressure p_c, the atmospheric pressure p_r, the area of the diaphragm A_d, the area of the supply plug A_s, of the exhaust plug A_e, and the spring force due to precompression F_s, the total force active in downward direction in case of a small increase in nozzle back pressure p_n is

$$p_n - [p_cA_d + (p_s - p_c)A_s + (p_c - p_r)A_e + F_s] \tag{6–3}$$

while in the case of a small decrease in nozzle back pressure, the forces in an upward direction are

$$-p_cA_d + p_n \tag{6–4}$$

Hence the nozzle back pressure may change by an amount equal to the difference between (6–4) and (6–3) or

$$(p_s - p_c)A_s + (p_c - p_r)A_e + F_s$$

without moving the diaphragm, i.e., without changing the controller output signal. A dead zone is thus created which can be minimized by design, but not eliminated.

The capacity of a non-bleed pilot valve is generally larger than that of a continuous-bleed pilot valve. A good design of the non-bleed type takes about 15 seconds to build up the pressure from 3 to 15 psi in a 350 cu in. volume. The continuous-bleed type may need 3 to 4 times that much.

Jet Pipes

The jet pipe is widely used with hydraulic control systems. Figure 6–17 is an illustration of jet pipe action in its conventional form. The jet pipe swings in a vertical plane around a horizontal axis, supported on its left by a ball bearing pivot, and on its right by a sleeve bearing through which a minute amount of oil is allowed to leak into the common sump (not shown). This method all but eliminates the friction in the suspension. The pressure at which the oil is pumped through the jet pipe depends on the application and is usually somewhere between 100 and 400 psi. It is kept constant at the predetermined pressure. From the jet pipe tip the oil is directed to the two closely adjacent receiving orifices in the distributor block, which are in turn connected with the opposite ends of the actuating cylinder.

With the jet pipe in its midposition, the same pressure exists in either

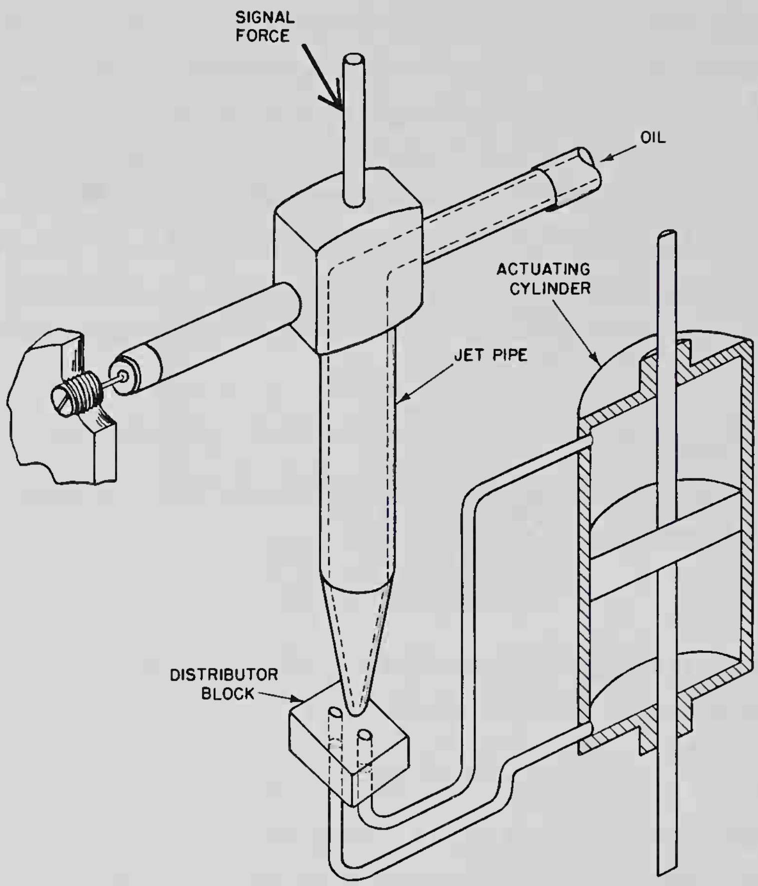


Figure 6–17. Schematic drawing of jet pipe action.

receiving orifice, consequently the pressures on either side of the actuating piston are equal. With the slightest deflection of the jet pipe, one orifice will receive more oil than the other, thus creating a difference in pressure between the two ends of the actuating cylinder and causing the piston to move in response to the jet pipe deflection.

As the oil leaves the converging jet pipe tip, its potential energy is changed into kinetic energy. As it enters the receiving orifices, kinetic energy is again changed into potential energy. This conversion of the oil pressure into velocity pressure in the jet stream between jet pipe tip and receiving orifice, and the subsequent reconversion into the static recovery pressure, results in a certain pressure loss. The recovery is generally between 85 and 96 per cent, which means that with a 100 psi supply pressure, the pressure at the actuating piston would be 85 to 96 psi when the jet pipe is fully deflected.

The relation between jet pipe displacement and static recovery pressure in the receiving orifices is shown in Figure 6–18. Under normal operation, the jet pipe deflection is limited by mechanical stops to about 0.035 in. on either side from center. The action is linear around the mid-position, though nonlinearities creep in toward the extremes.

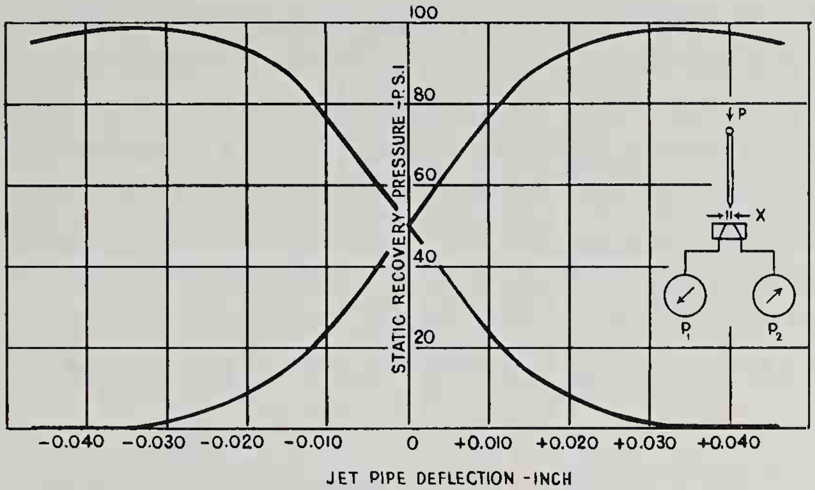


Figure 6–18. Static recovery pressure as function of jet pipe deflection.

Various jet pipe tips are available. Typical sizes and corresponding flows in cu in./min. at various supply pressures are as follows:

Jet Pipe Tip Diameter in Inch	100 psi	200 psi	300 psi	400 psi
0.050	115 cim	162 cim	200 cim	230 cim
0.065	194	275	336	388
0.080	294	415	510	588
0.100	460	650	800	920

Flow rates as shown are applicable when the jet pipe is fully deflected and the actuating piston does not deliver work. In the measure as load is applied to the piston, the load flow and hence the speed of the piston decrease. This is illustrated in Figure 6–19 which shows the decreasing load flow due to load pressure as well as to different deflections of the jet pipe.

One of the essential advantages of the jet pipe is that flow forces are practically non-existent, hence it can be operated with very weak input signals. On the other hand it is limited in its power output.

Four-way Valves

Where high power outputs are needed, the four-way valve is the most frequently used component in hydraulic circuits. It is frequently positioned by a jet pipe or a flapper-nozzle. If the input signal is strong enough, the spool of the four-way valve can also be positioned directly.

Figure 6–20 shows the basic concept of a four-way valve. In position

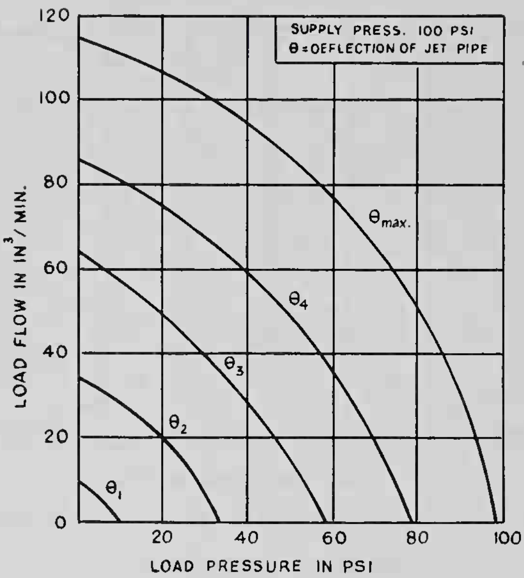


Figure 6–19. Typical pressure-flow characteristic of jet pipe.

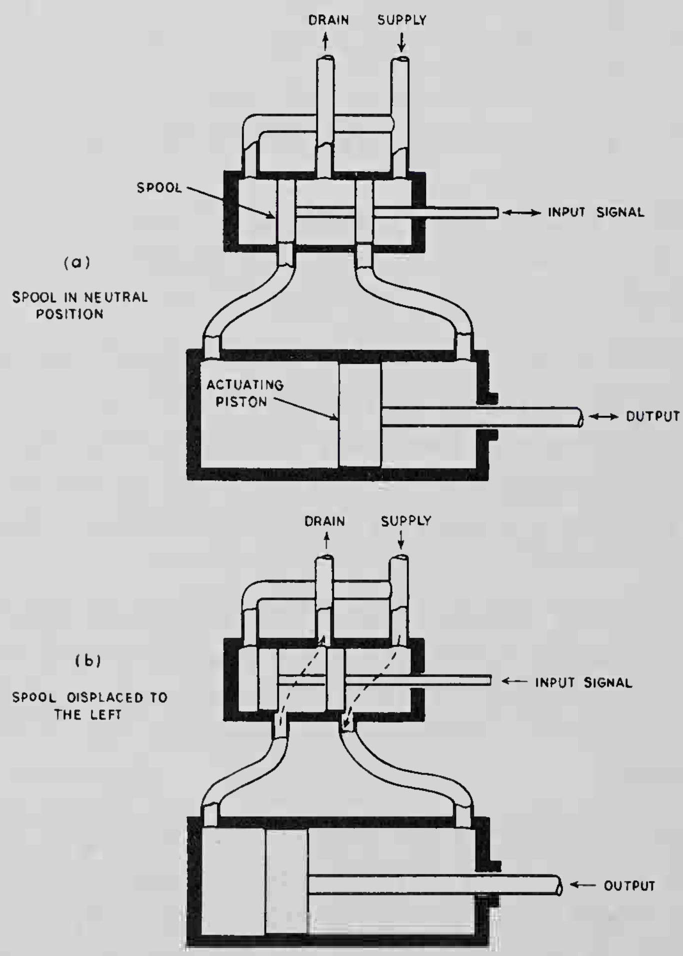


Figure 6–20. Schematic drawing of four-way valve.

a a certain hydraulic volume is trapped on both sides of the actuating piston. When the input signal displaces the spool to the left, as in position *b,* a connection between supply and the right side of the actuating cylinder is established, as well as a connection between left side and drain. The resulting pressure difference moves the piston to the left.

Ideally speaking, the spool in its neutral position should close the ports to the actuating cylinder and the slightest displacement should open them. Under manufacturing conditions this can hardly be obtained, and either overlap or underlap is the result as illustrated in Figure 6–21.

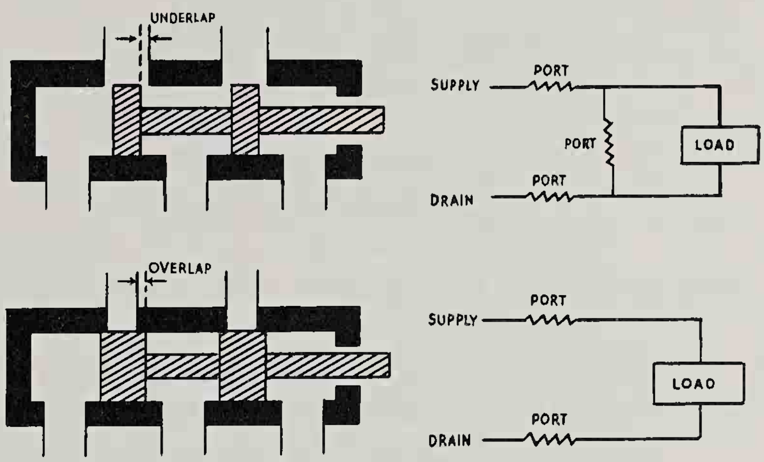


Figure 6–21. Underlap and overlap of four-way valve.

With overlap the lands on the spool are wider than the ports and there must be a certain displacement of the spool before the ports actually open. This approaches the conditions of a dead band which is obviously undesirable, and the underlapped condition is preferred.

The underlapped condition produces leakage flow around the valve ports. This results in a basic change in the behavior of the valve when load is applied to the actuating piston. With overlap and at a fixed piston position, the full hydraulic pressure is available across the load. The load sensitivity as illustrated in Figure 6–19 for the jet pipe does not exist for the jet pipe under these conditions. With underlapped conditions, however, a bypass around the load exists, and as long as the displacement is not larger than the underlap, this bypass produces a pressure drop in the supply port which increases with the load.

This condition is illustrated in the two graphs of Figure 6–22. These graphs were published by J. L. Shearer.* The symbol x designates the spool displacement. For a valve with zero underlap, corresponding to the

---

*Shearer, J. L.,"Dynamic Characteristics of Valve-Controlled Hydraulic Servomotors," *ASME Transactions,* (Aug. 1954).

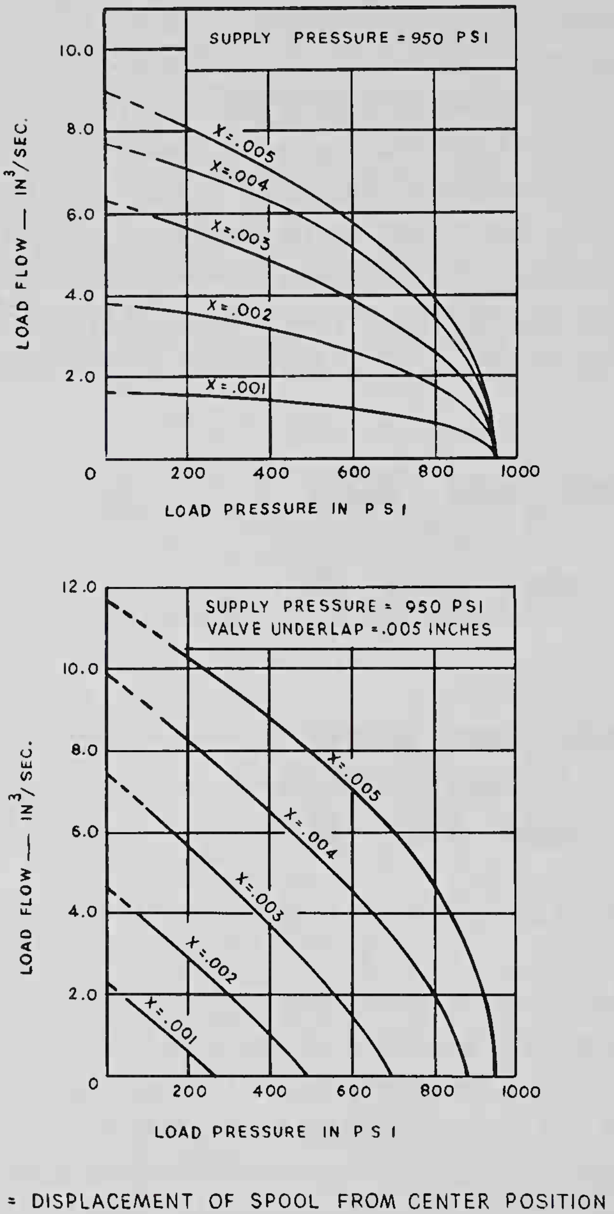


Figure 6–22. Pressure-flow characteristics of typical four-way valve.

upper diagram, the smallest displacement provides the full break-loose pressure of 950 psi across the actuating piston. With an underlap, as illustrated in the lower diagram, the characteristic changes completely. With a displacement of 0.001 in., the maximum obtainable pressure is only 250 psi.

The fluid that passes through the ports of the four-way valve causes flow forces along the spool axis. Figure 6–23 illustrates the flow through the valve ports. The flow enters at a certain pressure and velocity. At the metering port, i.e., at the outlet port, the velocity is greatest because the cross-sectional flow area is smallest. Hence the static pressure at the metering port is smallest. On the other hand, the velocity at the inlet

port is comparatively low and the static pressure is therefore high. The result is that the pressure components along the axis of the spool are larger toward the left than the right and tend to close the valve. Examination of Figure 6–23 shows that the flow may be reversed without alter-

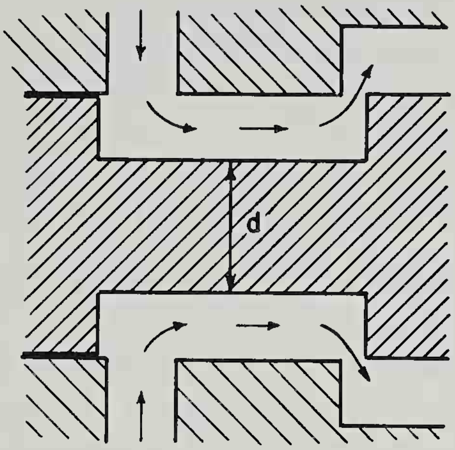


Figure 6–23. Flow through ports of four-way valve.

ing the direction of the flow force. The general statement holds therefore true that flow forces oppose the opening of such a valve. Their magnitude can be expressed within certain assumptions by

$$F = 0.000075Q\sqrt{p_1 - p_2} \tag{6–5}$$

where F is the flow force in lb, Q is the flow in cu in./min., and $(p_1 - p_2)$ is the pressure drop through the port in psi.

In a practical case, the pressure drop may be 900 psi, the flow 500 cu in./min., and the resulting flow force would be about 1⅛ lb.

Equation (6–5) is based on the assumption that diameter d (Figure 6–23) is small in comparison with the diameter D of the piston spool. By making d as large as practical, the velocity at the inlet port can be increased, and the flow forces will decrease correspondingly.

Special valve configurations have also been developed, such as the one shown in Figure 6–24, providing streamlined surfaces which cancel most of the flow forces. The relatively high cost of machining the necessary parts is the main disadvantage of such construction.

Hole-and-plug Valve. This is a special configuration of the four-way valve which has been developed at the Massachusetts Institute of Technology. The purpose of this design is to make a valve that approaches zero lap at a lower cost than is possible with the spool type discussed so far. As illustrated in Figure 6–25, the valve consists of a plate and a body, both made of rectangular stock and their flat, ground surfaces facing each other with a minimum of clearance.

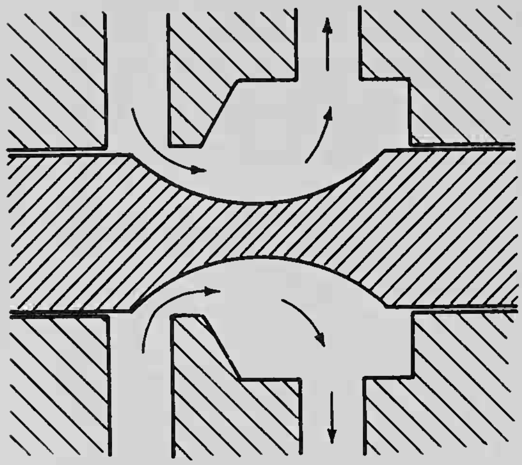


Figure 6–24. Flow through force-compensating ports of four-way valve.

Essentially, the operation is the same as in any four-way valve, i.e., it consists of a relative displacement of lands and ports. The difference in the two approaches lies in the production method. The blackened parts in Figure 6–25 are inserts. To prepare the holes into which they are inserted, the valve plate is aligned with the valve body and a hole is drilled straight through. This is considerably easier than making square ports as is customary in conventional spool valves to obtain linear flow versus displacement relation. Since the inserts in the hole-and-plug valve are round, the displacement of the plate and the corresponding opening of the circular hole result in linear relations.

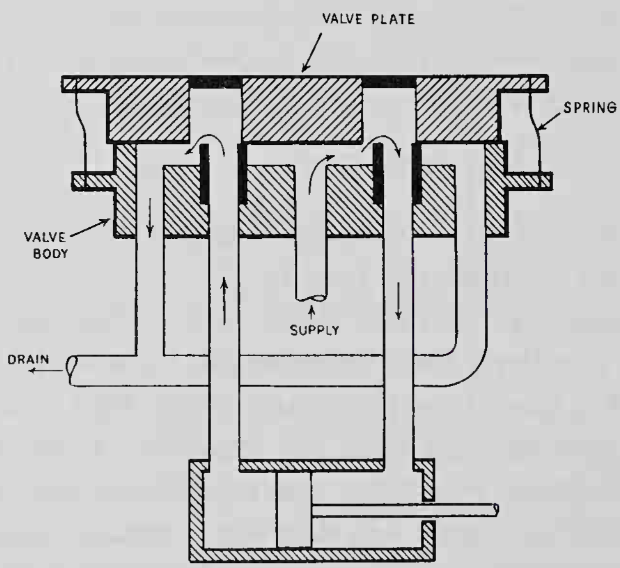


Figure 6–25. Hole-and-plug valve.

Assuming that the outer edge of the lower cylindrical insert is perfectly square, that the surfaces are flat and the interspace between plate and

body negligible, then a perfect zero-lap valve is obtained. Such ideal conditions can, however, only be approached but not attained. The result is a behavior that resembles that of an underlapped valve.

Cylinders

Figure 6–26a shows a single-acting cylinder, and Figure 6–26b, a double-acting cylinder. Flapper-nozzles are used for positioning in this example but other relays could be substituted. In the single-acting

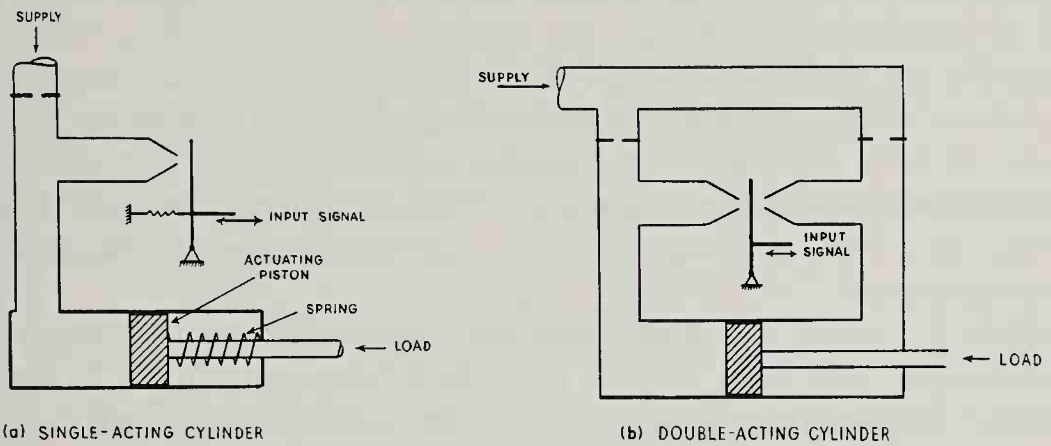


Figure 6–26. Single-acting and double-acting cylinders.

cylinder the nozzle back pressure is balanced by the spring in the actuating cylinder and the load on the piston. The disadvantage of this arrangement is the loss of energy spent in compressing the spring. An advantage, however, for a number of applications is the fact that a spring provides a safety feature against failure. In case of supply-pressure failure, the spring will drive the actuating piston to the left.

The double-acting cylinder requires a feedback arrangement in order to obtain a position proportional to the nozzle back pressure. The arrangement shown is that of a proportional-speed floating controller. The major advantage of the double-acting cylinder is that it operates without spring and that the full nozzle back pressure is available in any position to actuate against the load.

7. ELECTRICAL COMPONENTS

A rather large number of components are available to convert the magnitude of a controlled variable into an electrical signal. Input signals to these components are the *emf* of a thermocouple or of a pH electrode, the resistance change of a resistance temperature detector or of a strain gauge, etc. These signals are applied to an amplifier and its circuit or to a galvanometer. Diaphragms and bellows are other frequently used input signal systems which are then combined with a variety of electrical components to convert the mechanical into an electrical signal.

Figure 7–1 shows the most common of these components. They are by no means the only ones, but are those most frequently found in control systems. The following pages describe the components as listed in this illustration.

Potentiometers

The potentiometer consists essentially of a sliding contact and a slide-wire. The sliding contact is mechanically positioned, generally by rotary

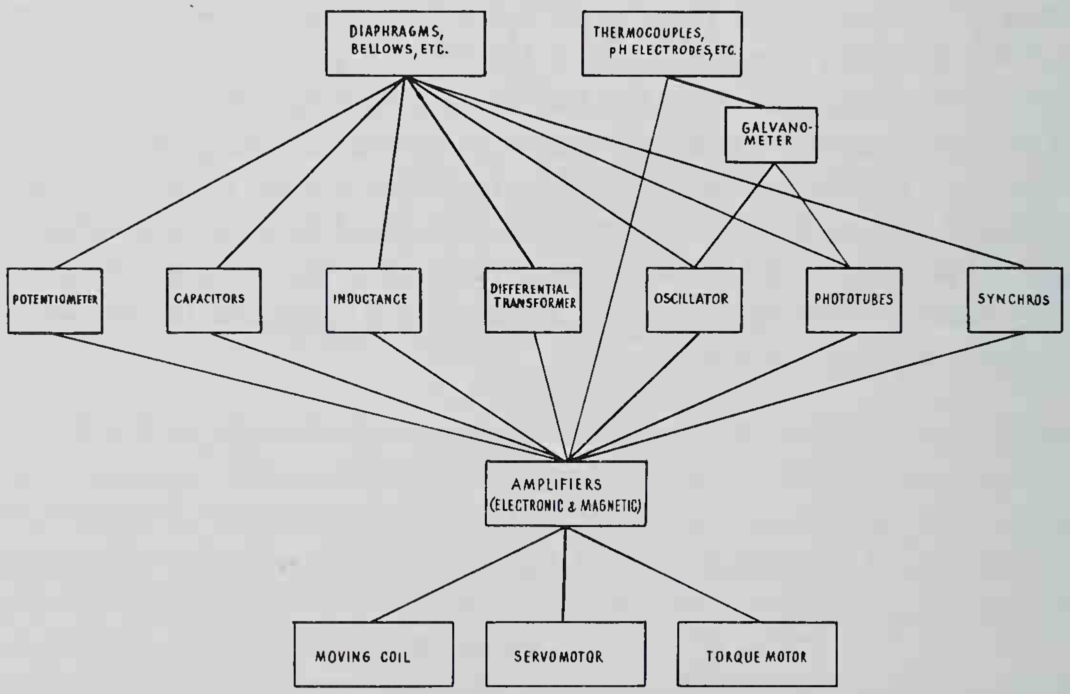


Figure 7–1. Components of electric control systems.

but sometimes by translational motion. Thus a device is obtained that can convert mechanical position into an electrical signal. Figure 7–2 shows the general arrangement. The supply voltage E_s is kept constant

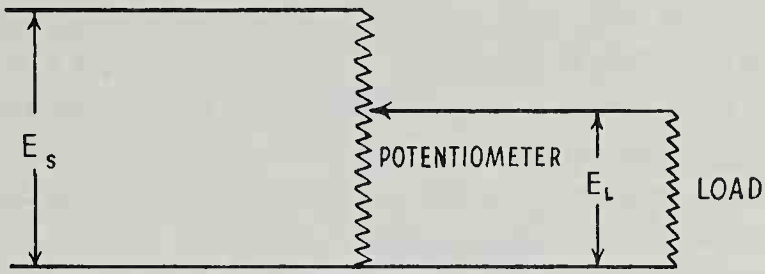


Figure 7–2. Potentiometer circuit.

and the load voltage E_L is varied by means of the sliding contact. The equation that expresses E_L is

$$E_L = \frac{kR_L}{kR_p + R_L - k^2R_p} \times E_s$$

where k defines the position of the sliding contact. R_L is the resistance of the load and R_p is the resistance of the potentiometer or slidewire. If all of the slidewire R_p is parallel with the load resistance R_L, then $k = 1$; if only half is parallel, then $k = 0.5$; etc.

The equation shows the nonlinearity of the relation between load and supply voltages. However, in making R_L large as compared with R_p, the equation approaches the linear relation $E_L = kE_s$.

The linearity behavior of the potentiometer in the circuit differs from that of the potentiometer as such. The latter refers to the relation of sliding contact position to resistance between sliding contact and slidewire terminals. As a rule this linearity is specified as normal or independent linearity (see p. 11).

The resolution sensitivity of a potentiometer is largely determined by the fact that the sliding contact passes from one slidewire turn to the next. The wire length of this turn represents a certain magnitude of resistance which is covered by the slidewire in one single step. Thus the resistance changes in steps rather than continuously. The finer the wire the more the turns and the smaller the steps, but at the same time, the finer wire represents a larger resistance, and the actual change in resistance is about the same. The best way to improve resolution sensitivity is to increase the sliding contact travel. If the travel is rotary, then an increase of radius will give best results. Multiple-turn potentiometers provide another method to increase resolution sensitivity. In either case the travel is increased and the resistance per turn is decreased. The total

resistance of the potentiometer remains the same, but the resolution sensitivity is improved.

The disadvantage of the stepwise response is overcome by using a single straight wire. This results generally in resistances which are too low to be practical. The wire may be replaced by conducting plastics or conducting films of high resistance. It can be expected that conducting plastics will in the future widen considerably the usefulness of potentiometers in control systems.

The resolution sensitivity of the potentiometer depends further on the static friction between sliding contact and wire, and in the bearing of the shaft which carries the sliding contact. A certain minimum torque on the shaft has to be developed before the sliding contact is loosened from the hold of this friction.

Frequently, the resolution—not the resolution sensitivity—of the potentiometer is specified. This resolution refers only to the stepwise arrangement of the windings, and not to the friction effects.

The contact between slider and wire constitutes in itself a resistance. This is produced by a film that develops by oxidation or other causes on the metallic surfaces. It causes local heating and the heat has frequently a cleansing effect although it may even worsen the conditions of the metallic surfaces. Resulting are minute changes in overall resistance which may be harmful in their total effect in control circuits. Vibration causing the contact pressure to vary may have similar effects. There are a number of other causes that may produce spurious changes in resistance. The resulting phenomenon is termed *noise.*

Most disadvantages of the potentiometer are due to the physical contact between sliding contact and slidewire, which produces friction and wear. A precision potentiometer has a life expectancy of about one million cycles. Considering that in a control system a sliding contact may easily move an average of 4 times per minute, and that this motion may concentrate on a limited portion of the slidewire, the life expectancy of such a potentiometer would be only one half year. This leads to the conclusions that the potentiometer should only be used where the sliding contact moves not more than occasionally, and that a noncontact device offers great advantages for all other applications.

Potentiometers and Rheostats

Figure 7–3 illustrates the difference between a potentiometer and rheostat. The potentiometer is an adjustable voltage divider, while the rheostat is an adjustable resistor which regulates current. Frequently, a choice between either one of the two elements has to be made. In control circuits, the potentiometers will be generally preferred. This is

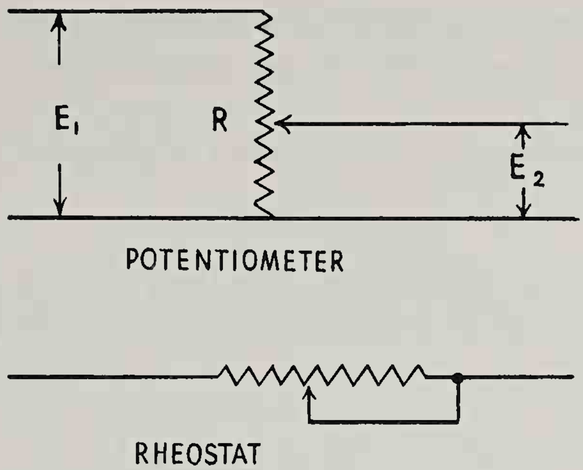


Figure 7–3. Potentiometer and rheostat.

because the accuracy of the output to input relationship of the potentiometer is only determined by the linearity of the winding resistance. In commercial precision potentiometers this is about 0.1 per cent. Commercial tolerances of total resistances are usually ±5 per cent. This has no influence on the overall accuracy of potentiometers since

$$\frac{E_1}{E_2} = \frac{R}{kR}$$

and variations of R do not alter the result. This is of particular importance where the magnitude of resistance is affected by temperature variations. In rheostats, a change of 5 per cent will change the current by the same percentage.

A further point to consider is that the potentiometer is connected across the power source and if the circuit through the load is open, there remains a shunt through the potentiometer. In the case of a rheostat the power supply is left with an open circuit when the load is removed.

Capacitors

The capacitor is an essentially frictionless device that changes the electrical characteristics of a circuit. Two metallic plates separated by air and connected into an a.c. circuit result in a capacitance which is expressed by

$$C = \frac{A}{4.5D} \times 10^{-12}$$

where C is the capacitance in farad, A is the minimum sq in. area of either plate which is directly opposite the other plate, and D is the distance between plates in inches. For example, a pressure-sensitive diaphragm can be arranged to deflect with the process pressure and at the

same time constitute one plate of a capacitor, while the other plate is fixed. Since the distance D changes with the deflection, the capacitance also changes.

The current through an electric circuit is given by

$$I = E/Z$$

where I is the current in amperes, E is the potential in volts, and Z is the impedance, i.e., the combination of capacitance and resistance, in ohms. Furthermore

$$Z = \sqrt{R^2 + \left(\frac{1}{6.28fC}\right)^2}$$

where f is the frequency in cycles per second and R is the resistance in ohms, as shown in Figure 7–4.

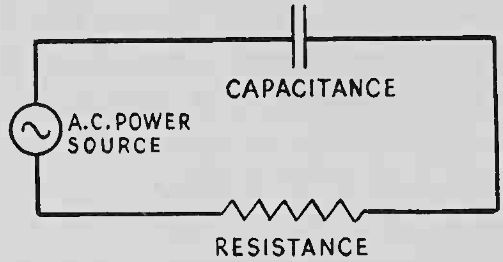


Figure 7–4. Resistance and capacitance circuit.

It is quite obvious from these equations, and can be verified by substituting numbers, that the capacitance has to be very large or the frequency very high to produce a current of reasonable magnitude.

Inductance Coils

Another resistive element in an electric circuit with alternating current is the inductance of a coil. When alternating current flows through a coil, a magnetic field is produced around the coil which pulsates in intensity and direction with the frequency of the current. This pulsating field generates an emf in the coil which opposes that of the circuit. This counter emf has effects similar to those of a resistor, and it is this effect that determines the inductance.

Figure 7–5 shows a circuit which combines resistance and inductance. The impedance of this circuit is

$$Z = \sqrt{R^2 + (6.28fL)^2}$$

where L is the inductance in henries.

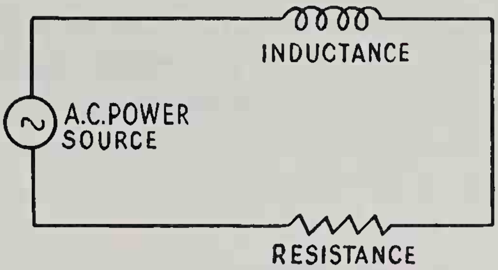


Figure 7–5. Resistance and inductance circuit.

If an insulated wire is wound about an iron core the inductance is given by

$$L = 376 \times 10^{-8} \frac{N^2 D^2}{l}$$

where L is the inductance in henries, N is the number of turns on the coil, D is the inside diameter of the coil in inches and l is the length of the iron core in inches.

If the iron core is removed the inductance becomes

$$L = 2.5 \times 10^{-8} \frac{N^2 D^2}{l}$$

In other words, the inductance changes by about 150 times when an iron core is introduced.

Since an iron core can be moved in a coil without physical contact, the inductance, and with it the impedance, of the circuit can be changed without friction.

Differential Transformer

Inductance as described above is the self-inductance of a circuit component. When two coils, each belonging to another circuit, are brought close together, mutual inductance results. This is the principle of any transformer. The alternating current in the primary winding produces a magnetic field within the vicinity of the primary which changes in intensity and direction with the frequency of the alternating current. The flux lines of this magnetic field cut the windings of the secondary coil. This induces an emf in the secondary. The magnitude of mutual inductance is expressed in henry, as is self-inductance. The mathematical relationships are also the same, and the difference between iron core and air core is similarly pronounced.

In the differential transformer, the effect of the iron core becomes further intensified because the secondary is split into two different coils which are arranged in various ways. One typical arrangement is illustrated in Figure 7–6. The two secondary windings are connected in such

a way that their polarities buck each other at any given moment. In the position shown, the iron core is in its center position, and due to the bucking effect the resulting emf in the secondary is zero. As the core is

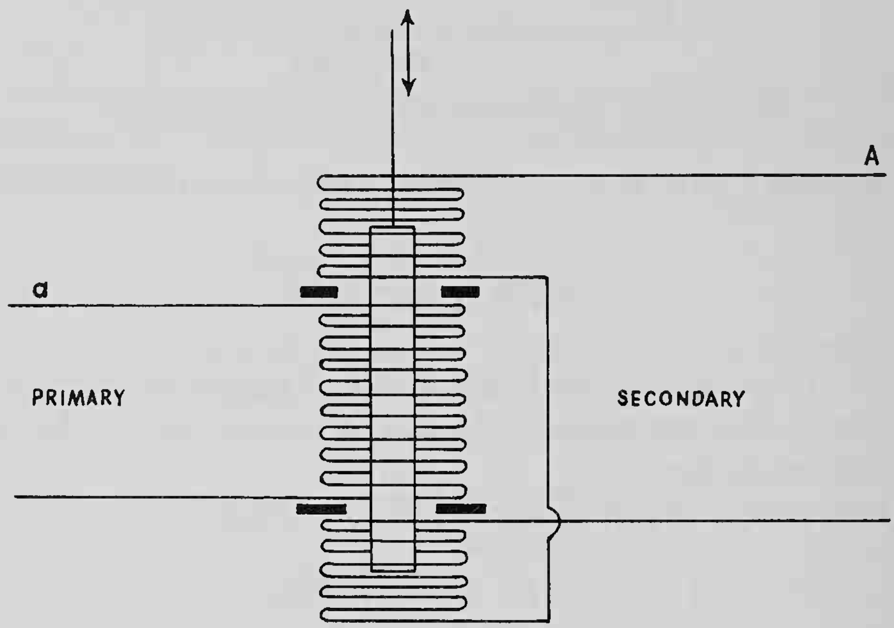


Figure 7–6. Differential transformer.

displaced upward, the emf in the upper winding increases, and simultaneously decreases in the lower winding. The result is an output equal to the difference between the emf's induced in the upper and the lower windings. Phase shift between the secondary and primary is caused by the magnetic circuit. Disregarding this phase shift, the polarity would be such that for any given instant when the current in the *a* wire of the primary has minus polarity, the current in the *A* wire of the secondary is also minus. If the core is displaced downward the output is equal to the difference between the emf's induced in the lower and upper winding, but phases are now reversed which means that for any given instant when the polarity of the *a* wire is minus, the polarity of the *A* wire is plus. The addition of phase shift between secondary and primary does not alter the basic phase reversal characteristic as described.

The differential transformer has perfect resolution sensitivity. Normal linearity is in the order of 0.05 per cent of maximum output over the specified stroke. The power of its output is low. This is due largely to the imperfect coupling between the primary and secondary and to the restriction in size that is generally imposed on the design of the differential transformer. As it stands, the output of differential transformers requires amplification to be utilized in control applications.

Galvanometers

The phenomenon of a magnetic field which is established by an alternating current flowing through a coil has been discussed under *Inductance coils*. If the alternating current is replaced by direct current, the magnetic field still exists. Since, however, it does not pulsate, no counter emf is produced and the current flow through the coil is determined only by its ohmic resistance. This steady magnetic field has a north and south pole and the intensity of the magnetic field is a function of the magnitude of the current flowing through the coil. If the coil is free to move, the coil will orient itself with respect to any other magnetic field that surrounds it because equal poles repel and unequal poles attract each other.

As seen in Figure 7–7, a small cylindrical airgap is left in a magnet circuit which is established by a permanent magnet with its pole pieces

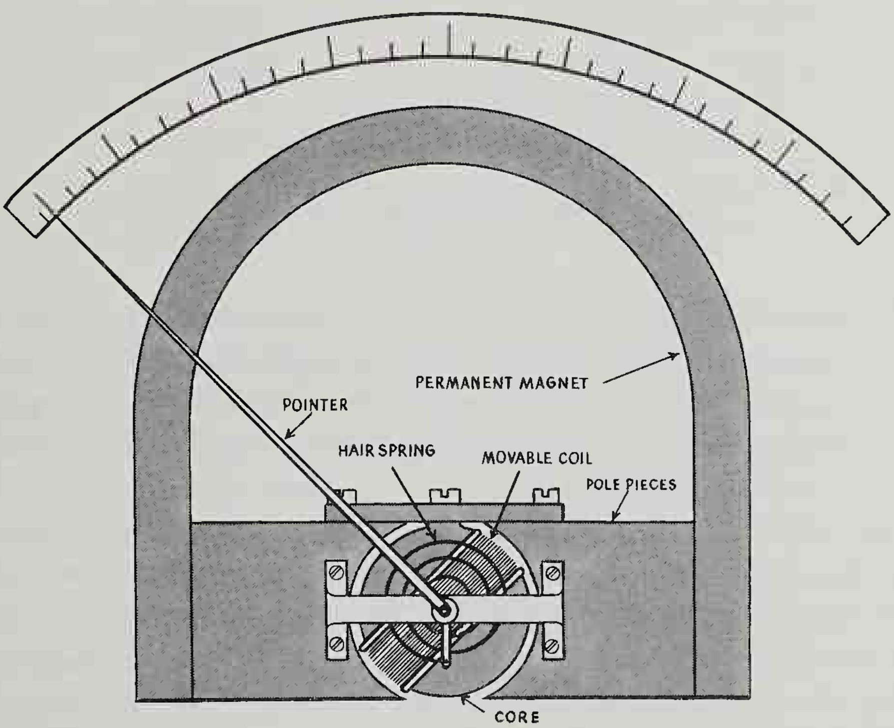


Figure 7–7. Galvanometer.

and a stationary core. With a d.c. current flowing through the movable coil, the magnetic field will interact with the flux of the permanent magnet. The resulting torque on the coil is restrained by the hair spring. Thus, the magnitude of the current flowing through the coil determines its deflection.

The torque of the galvanometer is comparatively small. The motion

of the pointer has to be picked up by other devices. Mechanisms that would periodically feel the position and relay this position into a more powerful action have been used for many years, but are now being gradually abandoned in favor of vanes mounted on the pointer moving between oscillation coils or interfering between a light source and a phototube.

Oscillators

The principle of mutual inductance is employed not only in the differential transformer, but also in the oscillator. Figure 7–8 shows a typical

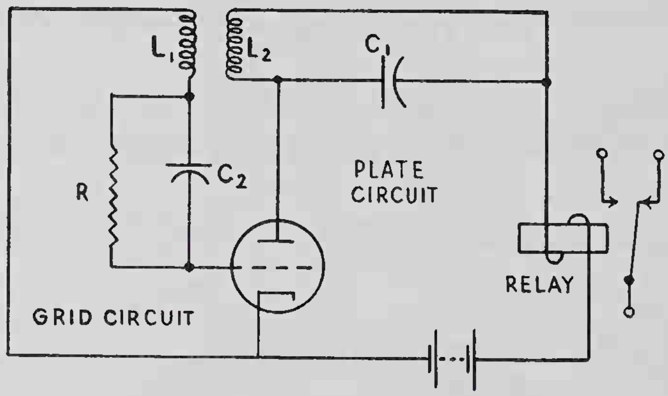


Figure 7–8. Oscillator control circuit.

oscillator control circuit. The plate circuit consists of the network combination of impedance L_2, relay coil, and capacitance C_1, in series with the power supply and plate and cathode of the tube. The grid circuit combines capacitor C_2, impedance L_1 and resistor R, which are in series with grid and plate of the tube.

Oscillations are produced because plate and grid circuit are coupled by impedances L_1 and L_2. As long as this coupling persists, a current will flow which keeps the relay energized. If a metal vane is inserted between the impedances, the vane acts as shield in the magnetic field, disturbing the mutual inductance. The result is that the current flow is interrupted and the relay deenergizes.

The combination of L_2 and C_1 is a so-called tank circuit. With the capacitor charged but no plate current flowing, the capacitor discharges into the inductance. The current flow induces a counter emf in the inductance which discharges into the capacitance. In case of pure inductances and capacitances this current pulsation could continue indefinitely. Since, however, some resistance is always present, the magnitude of this circulating current will rapidly decay, unless plate current begins to flow and supplies new energy.

When the vane prevents mutual inductance, the grid will be negatively charged because of the action in the grid bias circuit consisting of R and C_2. It is the principle of the grid controlled tube that a negative grid prevents flow of current in the plate circuit excepting the circulating current between C_1 and L_2. The grid bias circuit also contributes to limiting the amplitude as will be shown.

If the vane is removed, an even minute current flowing through L_2 will induce some voltage in L_1. The polarity of this voltage is such that it makes the grid less negative. This results in more current flowing through L_2 and more induced voltage in L_1, making the grid bias more positive and increasing the current through L_2, etc. The action is clearly a feedback action but with positive rather than negative feedback. The circuit is purposely made unstable. However, the instability, i.e., the amplitude of the oscillations, is controlled. This is largely due to the grid bias network. Suppose the voltage induced in L_1 is large enough to make the grid bias voltage positive for a part of the cycle. This results not only in an increase of the plate current, but also produces a considerable grid current. This charges capacitor C_2. As the grid becomes negative again, C_2 discharges through resistor R. The consequence is that the grid becomes more negative than it would without capacitor. This means that the more positive the grid may be at one moment the more negative will it be the next moment. This limits the plate current and hence the amplitude of the oscillations.

The oscillator circuit may also be so designed that the frequency, rather than the amplitude, of the oscillations changes with change in inductance. The principle is used in a Bristol controller.

Oscillators have the advantage of highest sensitivity. A vane motion of 0.002 in. suffices to change the oscillation in the circuit enough to energize or deenergize a relay. Since it is difficult to transmit high frequencies over any appreciable distances, it is therefore necessary to build the moving element with the vane and the complete oscillator in a single unit.

Phototubes

An ordinary vacuum tube emits electrons when the cathode is heated. The resulting flow of electrons is equivalent to an electric current, the plate current. The phototube differs insofar as the cathode does not need to be heated, but is light sensitive rather than heat sensitive. The current through a phototube depends on the voltage across it, the light intensity and also the light color.

There are three types of phototubes, namely, vacuum, gas-filled, and photomultiplier tubes. The current which passes through a vacuum

phototube at a given voltage potential is directly proportional to the light intensity. With the gas-filled phototube, the relation between current and light intensity is nonlinear. The gas-filled phototube has, however, the advantage of more change of current per unit change of light intensity.

Both tubes have generally an upper limit of 20 microamperes. The photomultiplier tube is usually rated for a maximum current of 20 milliamperes, i.e., one thousand times more than the other types. It is linear in its characteristics and has more change of current per unit change of light intensity than any other phototube. Gas-filled and vacuum tubes are, however, frequently preferred because of their lower cost. Besides, vacuum phototubes are of greater stability than any other.

Figure 7–9 shows a typical circuit as used with a phototube. The phototube is connected into the grid circuit of a triode. The grid current

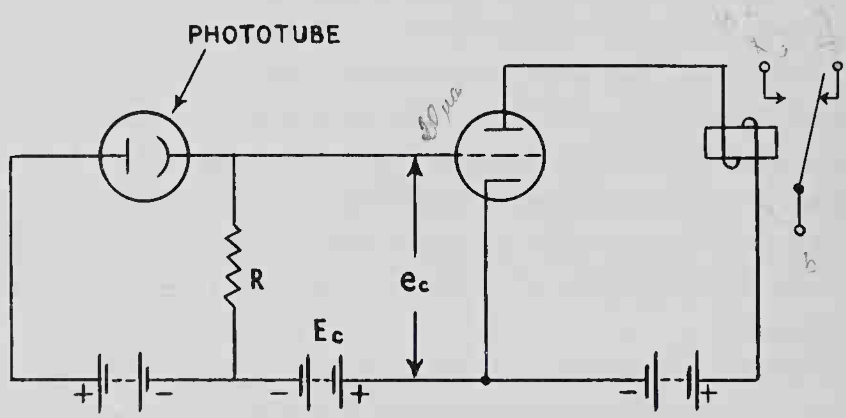


Figure 7–9. Phototube circuit.

is assumed to be zero. The current flowing through resistor R is then the same which flows through the phototube. The result is a voltage drop across the resistor with a polarity which is opposed to that of voltage E_c. For example, the phototube current is 10 microamperes and the resistance R is 2,000,000 ohms. The resulting voltage drop, $E = I \times R$, is 20 volts. If voltage E_c is 45 volts, then the grid voltage e_c is $20 - 45 = -25$ volts. The current through the phototube is determined by the light intensity. The current through the triode and the relay coil depends on the grid voltage. Hence the current through the relay becomes a function of the light impinging on the phototube.

Synchros

There are different synchro components, such as the synchro transmitter, synchro receiver, synchro control transformer, etc., which are combined in control circuits in various ways. In the following pages the

most important synchro components are described and typical combinations are shown.

Synchro Transmitter. The principle of a synchro transmitter is that of any transformer. An alternating current in the primary winding results in a magnetic field within the vicinity of the primary which changes in intensity and direction with the frequency of the alternating current. If a secondary winding is placed into the same magnetic field, a voltage is induced in the secondary and a current will flow through a load connected to it. This is the result of electromagnetic induction produced by an alternately building up and decaying of a magnetic field. If primary and secondary can be moved with respect to each other, then the maximum voltage is induced when the secondary is located so that a maximum flux from the primary passes through its winding. Depending on the direction of the flux lines with respect to the secondary, the polarity of the voltage in the secondary will be either in one direction or another.

The primary of the synchro transmitter is wound on a rotor, while the secondary is distributed in three separate windings on a stator as illustrated in Figure 7–10. Depending on the position of the rotor, more or

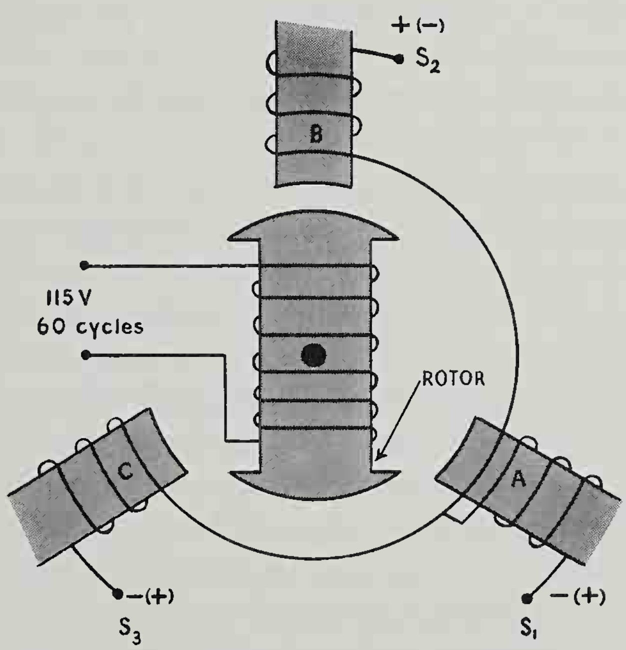


Figure 7–10. Synchro transmitter.

less of magnetic flux lines will cut through any one of the three windings. In the position shown, which is known as the zero position, voltages induced in windings *A* and *C* are equal and in the same direction. Hence the voltage between terminals S_1 and S_3 is zero.

The voltage induced in B is of opposite polarity to either winding A or C. Thus in a given moment, the terminal S_2 is of positive polarity, as indicated in Figure 7–10, while at the same time the polarities at S_1, as well as S_3, are negative. With the expanding and decaying magnetic field, polarities reverse, but as shown by an a.c. voltmeter, a voltage will be continuously indicated between terminals S_1 and S_2, as well as between terminals S_3 and S_2, but not between S_3 and S_1. Rotating the primary winding will change this condition gradually, since the number and direction of flux lines which cut through the secondary windings are altered. The result is illustrated in Figure 7–11. Each rotor position has

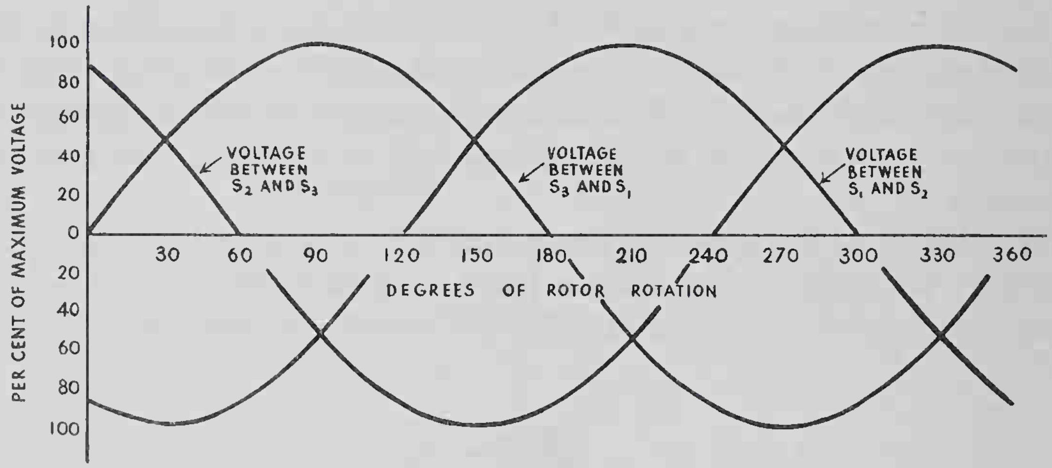


Figure 7–11. Output voltages vs. rotor position.

a corresponding pattern of voltages between the secondary windings. After rotating the primary 90 degrees from its neutral position, the voltage between S_3 and S_1 is the maximum obtainable in any position.

Synchro Receiver. The most basic synchro arrangement may be considered the combination of a synchro transmitter and a synchro receiver. The secondary voltages of the synchro transmitter are applied to the stator windings of the synchro receiver as shown in Figure 7–12. Essentially, the receiver is of the same construction as the transmitter. However, the input signal of the transmitter is a shaft position and its output is corresponding voltage. The input signal of the receiver is voltage and its output is position. The stator windings of the receiver set up a magnetic field in which the rotor, which is itself magnetized because of its windings, acts like the needle of a compass. The pattern of the field in the receiver follows the position of the rotor in the transmitter. Hence the receiver rotor—and with it the shaft—reproduces continually the position of the synchro transmitter shaft. The power which is available at the receiver shaft is small and suffices at best to position the needle of an indicator.

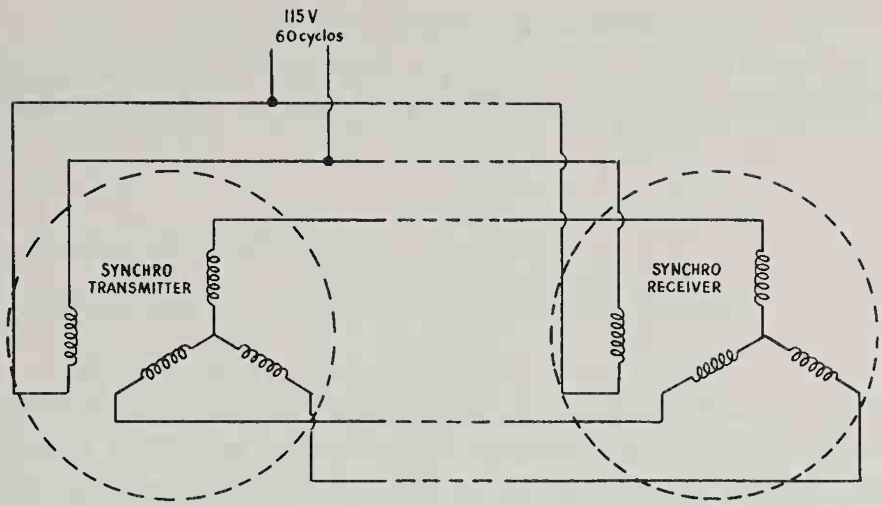


Figure 7–12. Schematic drawing of synchro transmitter and receiver.

Synchro Control Transformer. For closed-loop control purposes, the synchro receiver is replaced by the synchro control transformer. The control transformer is of essentially the same construction as either transmitter or receiver. The major difference is that the rotor shape is round, to prevent the magnetic field from producing a torque on the rotor. The only electric power connected to the control transformers is the signal voltage from the transmitter. This is applied to the rotor windings, while the stator windings represent the secondary of a transformer. The voltage induced in the secondary depends on the position of the rotor with respect to the magnetic field. For a field that corresponds to the zero position in the transmitter, the voltage induced in the secondary of the control transformer will be zero when the transformer rotor is displaced by 90 degrees from the transmitter rotor. This 90 degrees displacement applies for any other position of the rotor in the synchro transmitter.

Synchro Differential Transmitter. By arranging the rotor so that it carries three different windings distributed like those on the stator, a synchro differential transmitter is connected between the transmitter and the control transformer. The stator winding receives the signal from the transmitter. The output is modified by the position of the rotor. Thus the synchro differential transformer puts out a signal which is the sum of the shaft positions of the synchro transmitter and the synchro differential transmitter.

Combination of Synchros in a Control System. Figure 7–13 shows a typical arrangement of synchros. The controlled variable positions the rotor of the synchro transmitter. The output of the transmitter is connected to a synchro differential transmitter, where the set point is cranked in by positioning a rotor. The resulting actuating signal is

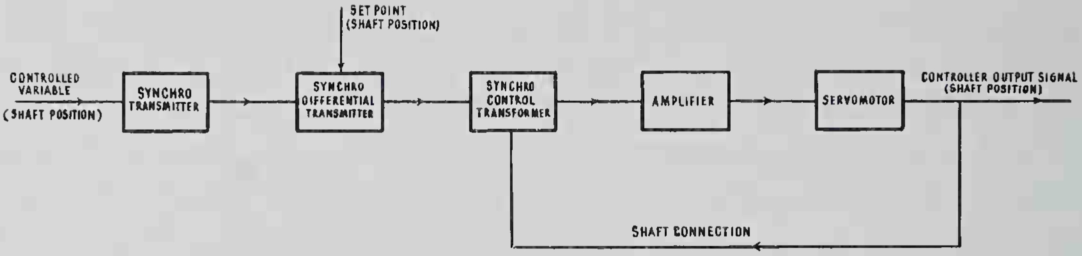


Figure 7–13. Synchros as part of control system.

applied to a synchro control transformer, the rotor of which is normally positioned so that the output signal to the amplifier is zero. However, it will deviate from zero when the actuating signal changes. This signal is amplified and transmitted to the servomotor, a device which will be described further below. The servomotor responds to the signal by rotation. Its shaft is linked with the rotor of the synchro control transformer. The rotation zeroes the output from the synchro control transformer, which in turn stops the servomotor. The same steady-state conditions as before are then obtained, but with a new position of the servomotor shaft. The shaft position which is the controller output signal can be used to position directly a final control element, or some intermediary element such as a slide wire potentiometer.

Other Usages of Synchros. The above descriptions do not intend to cover all modifications and applications of synchros. For example, a receiver similar to the differential transmitter may be built with triple stator windings. This permits applying to the stator as well as to the rotor signals which correspond with the angular displacements of two different shafts. The resulting rotor position of this differential receiver is then proportional to the sine of the difference between the angles of the other two shaft positions.

Another possibility is to obtain a primary feedback signal from small angular deviations, using for this purpose the synchro transmitter without any other synchros. Figure 7–11 showed that for limited deviations from zero, the voltage change is practically linear with the change in angle.

Electronic and Magnetic Amplifiers

Electronic amplifiers use either vacuum tubes or transistors or both. Gas-filled tubes—thyratrons—are used in the output stages when the required output power exceeds about 25 watts.

Transistors. Of the various forms of transistors—point-contact, junction, etc.—the junction transistor is used in the control field most frequently. The typical construction of a *p-n-p* transistor is illustrated

in Figure 7–14. It consists of a single crystal made from germanium or some other equivalent semiconductor. The *n*-type area conducts because excess electrons exist, while the *p*-type area has an electron deficiency

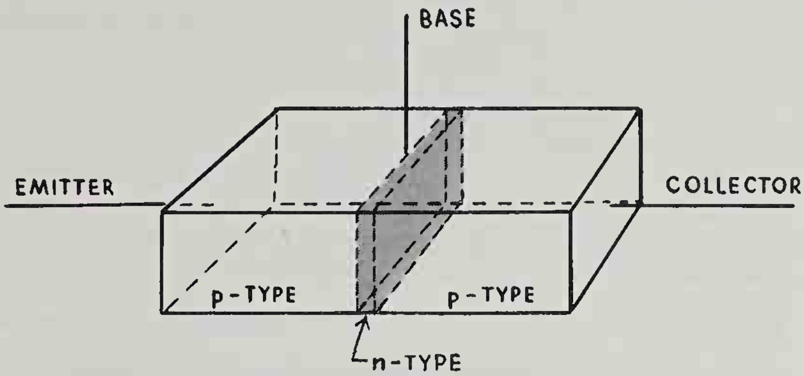


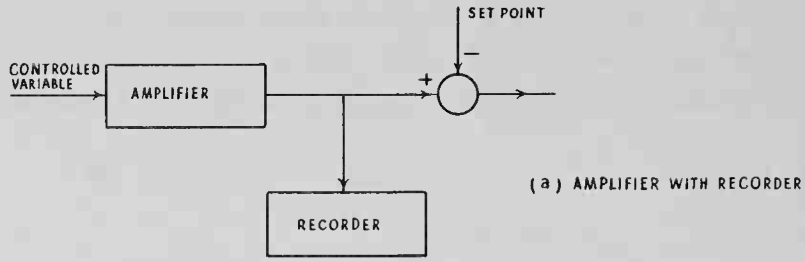
Figure 7–14. *p-n-p* Type junction transistor.

or so-called "holes," and conducts because electrons shift continuously to fill up the holes, only to produce holes where they leave. The center area which constitutes the base is very narrow, usually in the order of 0.001 inch. A *n-p-n* transistor differs from the *p-n-p* type in that the base is *p*-type, and emitter and collector are both *n*-type.

Transistors have replaced vacuum tubes in many electronic circuits. Their advantages are longer life expectancy, smaller size and less heat dissipation. The latter two characteristics are usually of lesser importance for industrial control systems. Some of the disadvantages however, are their temperature limitations, the variations of characteristics between transistors of the same type and manufacturer, and the voltage restrictions.

Ambient temperatures of 120°F are usually the operating limit for most transistors. In addition, the output changes with temperature and unless ambient conditions can be kept constant, the circuit design must consider these variations. Voltages in excess of about 30 volts can generally not be used. This compares with about 250 volts for vacuum tubes. The restriction is of particular importance where rate and reset actions are provided. Practically always, a resistor-capacitor combination is used for this purpose. The extent to which a capacitor can be charged depends on the voltage, and the reduced voltage of the transistor limits the usefulness of the resistor-capacitor network. Therefore, a vacuum tube will generally be used with rate and reset networks.

Electronic Amplifiers. Figure 7–15 illustrates two methods of using an amplifier. Accuracy requirements for the amplifier may depend on which method is used. In case *a* the amplifier output must be exactly proportional to the input since otherwise the recorder can not provide



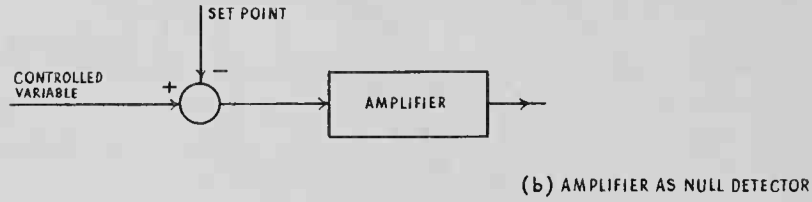


Figure 7–15. Different connections of amplifier.

a faithful reproduction of the controlled variable. In case *b,* however, the amplifier is a null detector. An output appears only when the controlled variable deviates from the set point. This allows relaxing the requirements for linearity and saturation. The term saturation means that the output signal changes with the input signal only to a certain point, at which the amplifier saturates and the output signal remains constant, though the input signal may change beyond the saturation point.

Drift and noise are characteristics which are disturbing in any application. For example, if the output of the amplifier starts to drift because of aging of parts, temperature changes, etc., a voltage will appear where zero is expected. Similarly, noise, which means spurious changes in output due to the thermal agitation of molecules in circuit components, or to electrons impinging on tube plates, may deviate the amplifier output from zero.

There is generally a preference of a.c. over d.c. amplifiers, because the latter have a greater tendency to drift. A d.c. input signal may be converted into a.c. by means of choppers, and then applied to a standard a.c. amplifier.

Magnetic Amplifiers. The magnetic amplifier is essentially a d.c. amplifier but, in general, does not have the tendency to drift like its electronic counterpart. It is by far the most rugged of all amplifiers. Since, however, even a vacuum-tube amplifier is hardly a fragile device, the higher initial cost of the magnetic amplifier is frequently its most serious handicap.

The ability of the magnetic amplifiers to convert small d.c. voltages into high-power output has led to their application in the temperature control of electric heat-treating furnaces. The power input to the heater elements of the furnace is directly regulated by the magnetic amplifier.

The d.c. input is from a thermocouple which measures the controlled variable.

Figure 7–16 illustrates the principle of the magnetic amplifier. The load current supplied from a.c. voltage passes through a rectifier. The

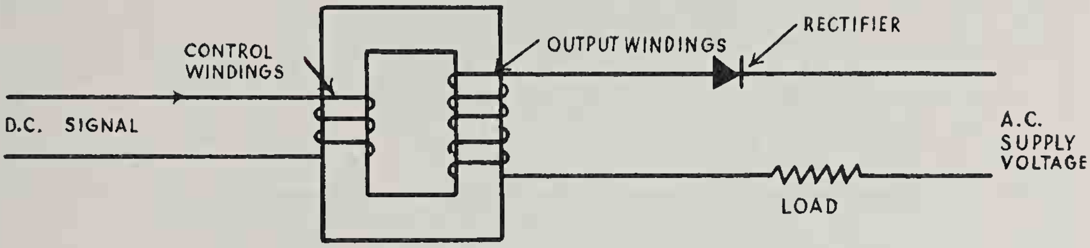


Figure 7–16. Principle of magnetic amplifier.

result is that current flows through load and output windings only during the positive half-cycles of the supply voltage.

The output windings act as an inductance in the load circuit. This inductance is, as stated before, the result of a changing magnetic field. Since the output windings are wound about an iron core, it is the magnetic field in the core that controls the inductance. The magnetization of iron can be increased only to its saturation point. Further increase of current does not change the magnetic field. In other words, once an iron core is saturated, the magnetic field remains constant, hence no counter e.m.f. is induced and the inductance is reduced to its minimum.

A magnetic amplifier is designed so that saturation is reached at approximately the positive peak of the supply voltage, while the control signal is zero. This means that before the peak is reached the inductance is at its maximum, hence only a minute current will flow. At the peak, however, the situation changes, the inductance vanishes, and the current obtains its maximum value. This voltage peak, however, does not coincide with the load current peak, since as in any inductive circuit the current lags behind the voltage. The saturation thus occurs before the actual current peak would be obtained.

The current flow keeps the core saturated. Hence, current will continue to flow throughout the remaining half cycle. After this, current flow is stopped because of the rectifier, and when the next positive half cycle starts, the same sequence as outlined above is repeated.

Signal current through the control winding either adds or subtracts to the saturation effect of the output winding. When the signal is positive, saturation—and hence current flow—appears before peak supply voltage is reached.

The time in which current flows can be shortened by making the sig-

nal negative. This is possible because the load current peak lags behind the supply voltage peak.

Thus the supplementary signal voltage controls the length of the current pulses through the load. This is illustrated in Figure 7–17 for various

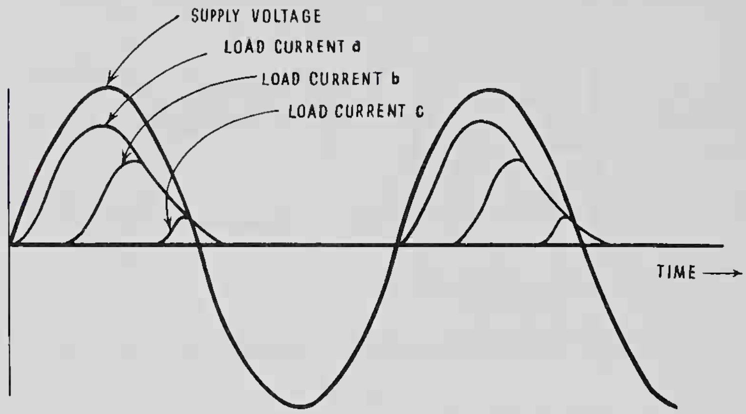


Figure 7–17. Load current at various signals.

values of load current. During the negative cycle of the supply voltage, current does not flow. This pulsating nature of the magnetic amplifier is overcome by the use of a bridge circuit combining two cores in a single unit, as shown in Figure 7–18. If *A* is positive, current passes through *B*

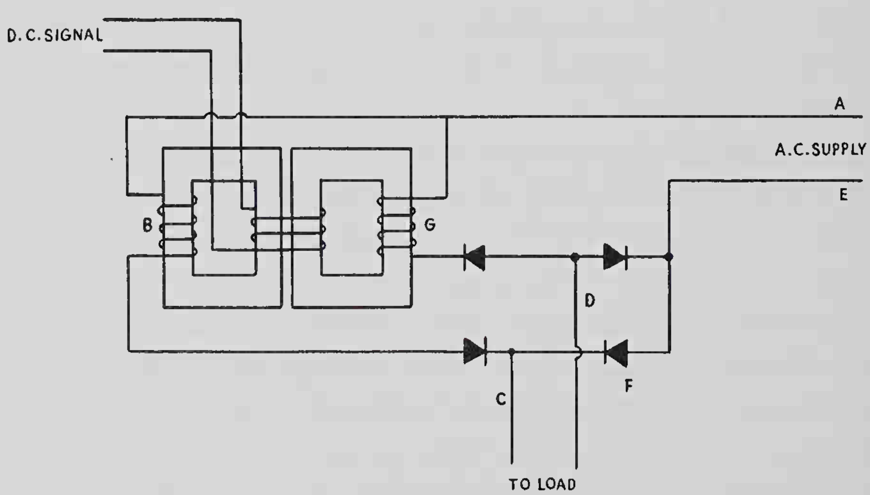


Figure 7–18. Full wave magnetic amplifier.

and *C* to the load, and back from there through *D* and *E*. As the phase reverses, and *E* becomes positive, current passes through *F* and *C* to the load, and back through *D* and *G* to *A*.

Moving Coil

Figure 7–19 is a schematic representation of the moving coil. It consists of a permanent ring magnet and an assembly made of iron to pro-

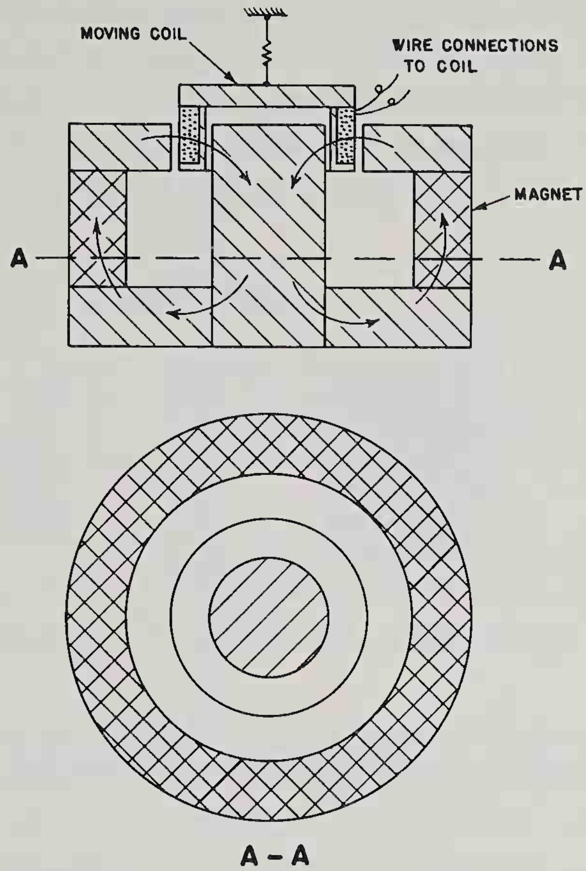


Figure 7–19. Schematic drawing of moving coil system.

vide a magnetic flux path as indicated by arrows in the illustration. The flux passes through an air gap where the moving coil is free to move. A signal current flowing through the coil results in a force of attraction or repulsion which is expressed by

$$F = 9 \times 10^{-9} BLi$$

where F is the force in ounces, B is the magnetic flux density in the air gap, in gauss, L is the length of wire that makes up the coil, in inches, and i is the current in milliampere.

The above equality shows that even a weak current could be made to produce a force by increasing either the flux density or the wire length, or both. This of course has practical limitations. For example, 4 milliamperes d.c. is a common output signal range of electronic controllers. For a force of 10 oz. the product of flux density and wire length would have to be

$$BL = 2.8 \times 10^{8}$$

Even with a coil that has 5000 ft. of wire the required flux density is still some 4700 gauss. Considering an air gap of 2 in. mean diameter and 0.5

in. height, the required cross-sectional area of an Alnico V ring magnet is about 4 sq in. This illustrates that bulk becomes a limiting factor when the force available from a moving coil is considered.

The permanent magnet can be replaced by an electromagnet. This is illustrated in Figure 7–20. There are two windings, one for the moving

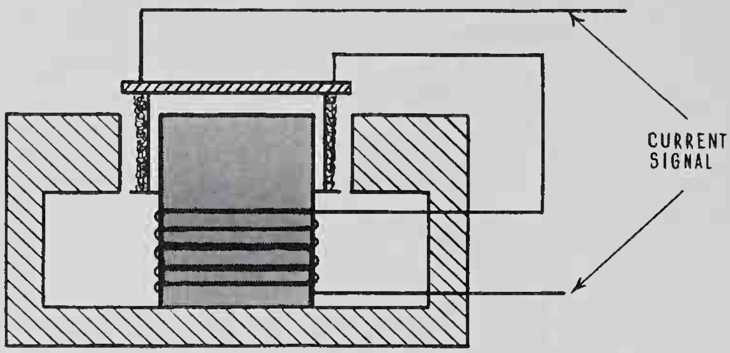


Figure 7–20. Electromagnet moving coil.

coil, the other for the electromagnet. The resulting force is now proportional to the product of the currents in the two windings. Since they are connected in series, the current in both windings is the same and the force is proportional to the square of the current. An important computing element is thus obtained. Its application will be described in Chapter 9.

Servomotors

Figure 7–21 illustrates the principle of operation of a two-phase induction motor. The rotor is represented by an iron bar. No electrical connection exists between the rotor and the stationary poles and windings. The windings are connected to a two-phase a.c. electric power supply, so that *a* and *b* are connected in series with one phase and *c* and *d* in series with the other phase. The windings are arranged so that with the direction of current flow as assumed in the illustration, there will be a north pole at *a* and a south pole at *b*.

The phase angle between the phases of a two-phase system is 90 degrees, which means that when the current in phase 1 is maximum, the current in phase 2 is zero, and vice versa. As the phase of the power supply proceeds, pole *d* will gradually magnetize and become a north pole, while the current magnetization in *a* decays. After this pole *b* will become north and so on. A revolving field is thus established and the iron rod will rotate with it.

If either phase were to be cut off, the torque available from the rotor would be correspondingly reduced. Furthermore, the motor could not start by itself.

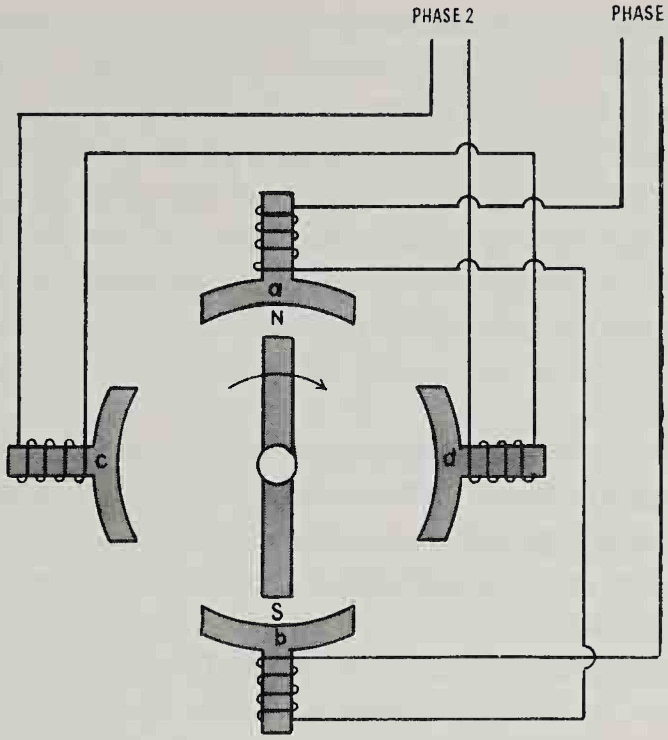


Figure 7–21. Schematic drawing of two-phase induction motor.

A servomotor is generally a two-phase induction motor. The so-called reference voltage obtained directly from the line is applied across windings *a* and *b,* while windings *c* and *d* are connected across the control voltage from a servo amplifier. The control voltage is reversible in phase, i.e., it either leads or lags the reference voltage by 90 degrees. This determines the direction of rotation.

The magnitude of control voltage depends on the actuating signal fed into the servo amplifier. The torque on the servomotor shaft is approximately proportional to the magnitude of the control voltage.

The rotor, shown as an iron rod in Figure 7–21 for the purpose of illustration, is replaced in an actual servomotor either by a squirrel-cage, solid cylindrical or drag-cup rotor. The important consideration is that the rotor be of light weight in order to reduce its inertia as much as possible. This permits it to obtain speed rapidly, and to be easily stopped and reversed.

In general the servo amplifier will combine an a.c. and a d.c. output signal. The d.c. signal is constant and is present even when the a.c. component is zero. This keeps the corresponding poles at constant polarity, providing an efficient braking action.

A significant difference between a servomotor and a conventional two-phase induction motor is that the latter will rotate when power is applied, while the servomotor will frequently stall though the reference

winding is energized. This produces considerable heating. As a result servomotors can be used with limited power output only, from 0.5 to 100 watts. Where the power output is of 25 watts or more, a blower is usually needed for cooling. Stalling torques from 0.5 to 95 oz/in. are available. This value can be increased by gears, since the high rotational speed is generally less a requirement than the torque output. Gears, however, must be used with caution, since backlash can readily cause instability of an entire control system.

Torque Motors

Restraining the motion of a servomotor by a spring allows it to move through a limited angle only. When the torque due to control voltage equals the force of the spring, the motion stops. Thus the angle is roughly proportional to the control voltage. A servomotor used in this form is called a torque motor.

Another type of torque motor is illustrated in Figure 7–22. Though the designation *motor* is hardly justified, it is this device which is more fre-

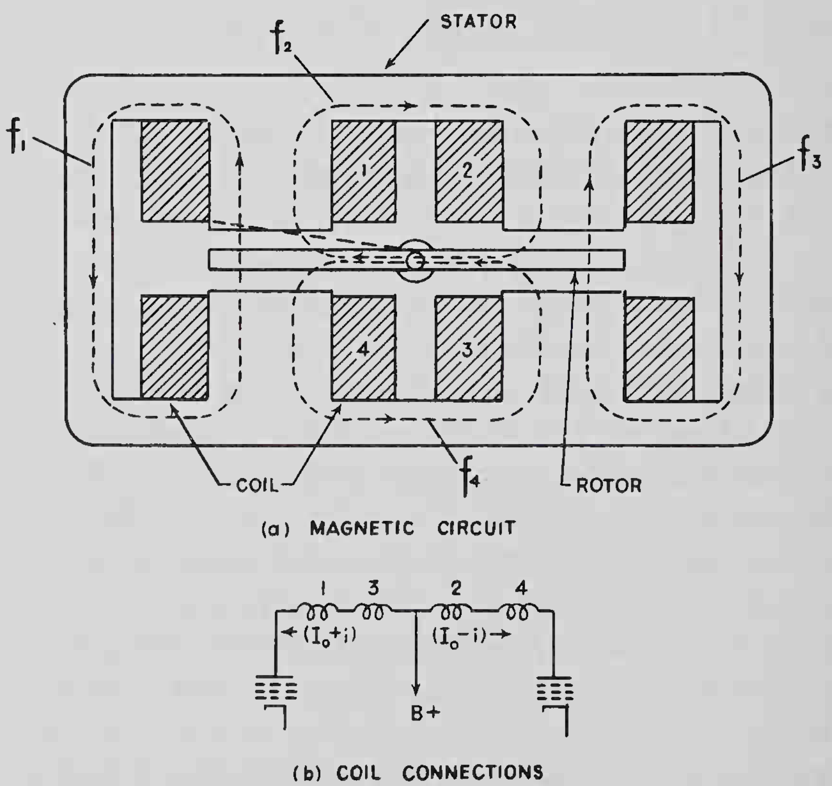


Figure 7–22. Magnetic circuit and coil connections of torque motor.

quently referred to as torque motor than the spring-restrained servomotor. This torque motor is basically a high-speed polarized type of relay. Coils 1 and 3 in the illustration are connected in series on one side of a push-pull amplifier, and coils 2 and 4 are in series on the other

side. The current flowing through these coils consists of a quiescent current I_0 and the signal current i.

When the signal current is zero, magnetic fluxes f_1, f_2, f_3, and f_4 are equal and the resulting torque on the rotor is zero. However, if the signal current departs from zero it is added in coils 1 and 3 and subtracted in 2 and 4, or vice versa. In this case fluxes f_1 and f_3 remain about the same, but f_2 and f_4 change, so that the magnetic flux in one pair of diagonally opposite air gaps increases while it decreases in the other pair. This produces a torque on the spring-restrained armature. The resulting torque is proportional to the signal, but to obtain this proportionality the armature motion must be limited to a very small angle.

The comparatively high hysteresis of these torque motors, which amounts to two or three per cent, is generally acceptable in control circuits. A somewhat more serious consideration is the differential connection as illustrated. Commercial electronic controllers do not generally provide this kind of signal, and an additional amplifier may therefore be required.

8. BALANCES AND COMPUTING CIRCUITS

Measuring means, controlling means and final control elements have in common elements, which convert input signals into output signals. The relation of output to input may be proportional, inversely proportional, a function of time, etc. If several inputs are combined, the output may be the sum, the ratio or some other mathematical relationship. This chapter will discuss principles and examples of proportional elements (implying in this expression also inversely proportional elements), and computing elements which combine several inputs. Those of specific control function, particularly reset and rate elements, will be described in Chapter 10. One other class, namely those elements which produce an output proportional to the square root of the input, is discussed also in Chapter 9.

Figure 8–1 shows a block representing a proportional element. Within this block a transformation takes place that changes an input into an

Figure 8–1. Function of a proportional element.

output signal. Some kind of balance system produces this transformation. Balances are specific forms of feedback arrangements. They are divided into force balances and position balances. A force balance operates either as deflection or a null system.

The force balance is well represented in weighing scales. A beam scale indicates the weight when the deflection is zero. This is a null system. In the spring scale the deflection is proportional to the force. This is a deflection system.

The deflection system always involves a spring. This is quite obvious when Hooke's law, $F = kx$, is applied. In this equation, F is the force exerted by a deflected spring, k is the spring rate, a constant, and x is the spring deflection. If W is the weight of the suspended object, then $W = F$, which is the expression of the null system; but it is also true that $W = kx$, which is a deflection system, because x is a deflection.

Since, however, $W = kx$, this also means that the deflection system predicates a spring because of the factor k which stands for the spring rate.

The equilibrium of a position balance is obtained by displacement and not by forces as in the force balance. A typical example of a position balance is the flapper-and-nozzle arrangement of a pneumatic controller as illustrated in Figure 8–2. The flapper is displaced to obtain a balanced

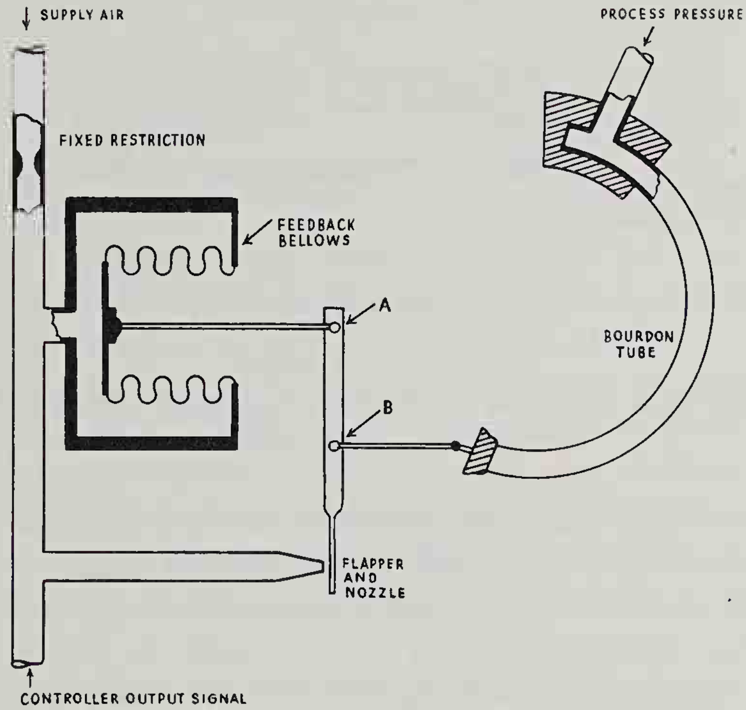


Figure 8–2. Flapper-and-nozzle system.

position in responding to two counteracting motions; one from the measuring signal, the other from the output signal. The flapper is not submitted to strain at any moment. When the bourdon tube responds to a change in controlled variable, it rotates the flapper around pivot A. This changes the nozzle back pressure and hence the controller output signal. The feedback bellows responds to this change and rotates the flapper around pivot B in such a direction that the initial clearance between flapper and nozzle is approached.

Force Balances with Deflection Systems

A typical example of a force balance deflection system is the galvanometer described in the previous chapter. The torque of the coil is produced by the current flowing through it. The motion is restrained by the hair spring. As the coil rotates, the counterforce of the hair spring changes until an equilibrium between the two forces is obtained. This is

a force balance. The deflection system is characterized by the fact that deflection is required to obtain equilibrium.

Another application of a deflection system in force balances is found in pneumatic systems, e.g., in the pressure-duplicator method, frequently used in converting liquid level into an equivalent air pressure. Figure 8–3 illustrates the principle involved. The top side of the diaphragm is

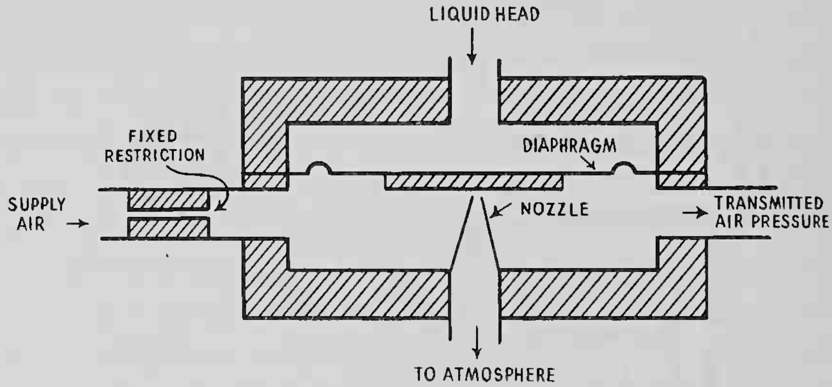


Figure 8–3. Pressure duplicator—pneumatic force balance with deflection system.

exposed to the liquid head. Air pressure is admitted to the bottom side, through a fixed restriction. The bottom chamber communicates directly with the receiver. A nozzle opening decreases when the diaphragm moves toward it. The further open the nozzle outlet is, the more air bleeds through the nozzle to the atmosphere and the lower is the pressure in the chamber. Suppose the liquid level rises, exerting an additional pressure on top of the diaphragm. It is assumed that the air supply pressure is higher than the maximum pressure due to liquid level head. The diaphragm will then tend to close the nozzle which will make the pressure in the bottom chamber rise until a balance is reached where the forces on the top and bottom of the diaphragm are equal. The arrangement is therefore a force balance, and it is a deflection system in that the new air pressure requires that a certain deflection of the diaphragm be maintained to partially close the nozzle opening.

Force Balances with Null System

The galvanometer is widely used in pyrometers—in millivoltmeter types and in some potentiometer types. When a potentiometer uses a galvanometer, the latter no longer exercises the function of quantitative measurement. Figure 8–4 shows the typical potentiometer circuit. There are two sources of emf: the battery and the thermocouple. Their respective circuits have one portion of the slidewire in common. The relative length of this common portion can be either increased or decreased

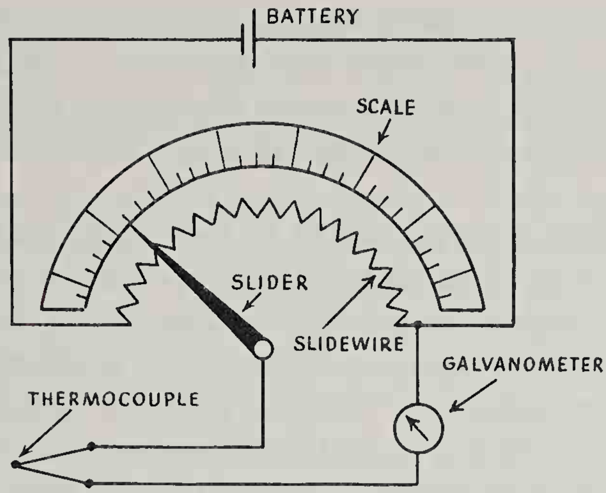


Figure 8–4. Potentiometer—electric force balance with null system.

by means of the slider. In balanced conditions no current flows through the thermocouple because of the bucking effect of the two emf's in the common slidewire portion. If the thermocouple emf changes, it is necessary only to readjust the slider in order to reduce to zero any current flow due to this new condition. Since a scale can be made to indicate the position of the slider, it can be calibrated in temperature units which express the reading of the thermocouple. The sole function of the galvanometer is to detect an unbalance in the circuit, but not the magnitude of the change.

The balancing action in the potentiometer can be compared with the beam scale mentioned before. One dish of the scale carries an object of unknown weight, the "input," and known weights are added in the opposite dish until the deflection of the balance is zero. The magnitude of the known weights that were added expresses the measurement, i.e., the "output." The input of the potentiometer is the emf of a thermocouple which produces a deflection of the movable coil in the galvanometer. Balance will be obtained by a known emf—in lieu of weights—which is mechanically adjusted until the coil deflection is zero. The amount of mechanical adjustment required is read in terms of the variable measured by the thermocouple.

As far as force balance is concerned, the potentiometer represents it as much as the millivoltmeter. The potentiometer, however, does not require the deflection of a spring in order to obtain balanced conditions.

It is an outstanding characteristic of the null system, as compared with deflection systems, that the former requires an outside medium activated by some added power source to make it work, while the latter can be self-contained and respond on its own accord. The added power source may be the emf of a battery, weights, air pressure, etc., and the

outside medium may be either a human operator or some automatically operated mechanism, commonly called feedback. Since the deflection system does not require such additional provisions, it appears that it has the advantage of greater simplicity. The null system does have advantages, however.

The beam balance always returns to the same balanced conditions. At the time when it indicates the measurement, there are no mechanical deflections in the system. In the spring scale this is different: for each measurement there is another configuration of mechanical parts. These movements must be free from hysteresis and nonlinearities—and in this respect there is a limitation in any spring. A further limitation is the influence of temperature changes on spring members, which may produce inaccuracies in their response to changing loads.

The null system has the advantage of not depending upon the deflection of spring members. It also allows greater sensitivity. When a millivoltmeter, i.e., a deflection system, is to operate over a certain range, e.g., 0 to 500°F, the movable coil will have its maximum deflection at 500°F, and any smaller values will produce correspondingly smaller deflections. It is quite feasible that the galvanometer in a potentiometer, i.e., a null system device, may obtain full deflection at, say, 10°F, provided that restraining members will allow the stress, though not the movement, for sudden changes beyond that range. The response of the galvanometer in a potentiometer to a given change can thus be made much greater than in a millivoltmeter. A smaller change of temperature can therefore be sensed. However, this assumes at the same time that the dead band in the system is negligible. If the system is unable to respond to a change of 1°F, neither the null system nor deflection system will improve the situation.

Since the galvanometer when used in a potentiometer is part of a closed loop and hence reduced to the function of an unbalance detector, it becomes of relatively little importance whether or not its response is linear. It is only necessary that it detects the unbalance, while the determination of the magnitude is obtained by auxiliary means.

The Null System in Pneumatic Circuits

This characteristic of permitting nonlinearities makes the null system particularly desirable in pneumatic flapper-and-nozzle circuits because the degree of nozzle opening is not entirely proportional to the transmitted pressure.

The pressure duplicator in Figure 8–3 operates very well as long as the spring constant of the diaphragm is low enough to be ignored. If this is not the case, the deflection of the diaphragm which is necessary to change the nozzle opening results in an added force and the trans-

mitted air pressure is no longer proportional to the pressure of the liquid head. For a metallic diaphragm, for example, the spring constant may be quite high and the accuracy of the device would suffer. One method of reducing this error would be to keep the required motion of the diaphragm as small as possible, i.e., to approach a null system.

While this can be done by making the nozzle diameter relatively large as compared with the motion of the diaphragm, or by reducing the size of the fixed restriction, one may also apply the principle of the pneumatic null balance, which was introduced by C. B. Moore. This uses an ingenious feedback principle which calls for maintenance of a relatively constant pressure drop across the nozzle and thereby reduces the deflection of the diaphragm considerably.

In order to clarify what is involved in converting the arrangement of Figure 8–3 into a null balance, equation (6–1) from Chapter 6 is repeated here, which equally applies for Figure 8–3:

$$A_1\sqrt{(p_s - p_t)p_t} = A_2\sqrt{(p_t - p_r)p_r} \tag{8–1}$$

where A_1 is the flow area of the fixed resistance, A_2 the flow area of the nozzle, p_s is the supply pressure, p_t is the transmitted pressure, and p_r is the reference pressure which in Figure 8–3 is the atmosphere. All pressures are absolute pressures.

In maintaining the diaphragm position, A_1 and A_2 are to be constants and may be combined into a single constant K, so that

$$K = \left(\frac{A_2}{A_1}\right)^2$$

This can be substituted in equation (8–1) and makes

$$K = \frac{(p_s - p_t)p_t}{(p_t - p_r)p_r} \tag{8–2}$$

In this equation p_t, the transmitted pressure, is necessarily a variable. In order to make the expression constant as a whole, both p_s and p_r would have to be varied in direct proportion to p_t, so that

$$p_r = ap_t$$
$$p_s = bp_t$$

where a and b are proportionality constants. Substituting this in equation (8–2) gives

$$K = \frac{(bp_t - p_t)p_t}{(p_t - ap_t)ap_t} = \frac{b - 1}{a(1 - a)}$$

thus eliminating all variables, and obtaining a null system.

The methods, however, of changing p_s and p_r in proportion to p_t are imperfect and complicated. Nevertheless the ideal null system can be closely approached. In view of the mechanical complexities, however, it is frequently preferred to leave the supply pressure constant and vary only the reference pressure. Even under these conditions it is possible to reduce considerably the diaphragm deflection.

Kirk* gives figures in one particular case for the diaphragm deflection required to change the transmitted pressure from 3 to 18 psig. For constant supply and reference pressures the diaphragm motion was 0.00209 in. By changing the reference pressure in proportion to the transmitted pressure, this motion was reduced to 0.00171 in. By changing both, the reference and the supply pressure, in proportion to the transmitted pressure, the motion was only 0.00011 in.

The principle which is used in a large variety of components of the Moore Products Company and which employs a variable reference pressure is illustrated schematically in Figure 8–5. It combines the pres-

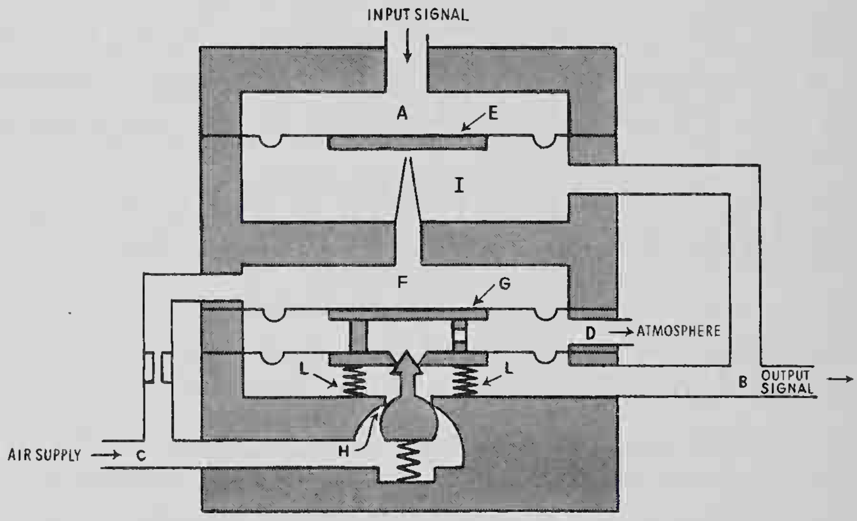


Figure 8–5. Pilot relay with variable reference pressure.

sure duplicator of Figure 8–3 with a pneumatic relay to increase the flow capacity for the transmitted pressure. The input pressure A is converted into an output pressure B by either admitting air pressure from the supply line C or by releasing it from B through D to the atmosphere.

When the input signal increases, diaphragm E moves downward, decreasing the nozzle opening and thereby increasing the nozzle back pressure F. The diaphragm assembly G, consisting of two diaphragms rigidly connected together, moves downward in response to the increased

* Kirk, D. B., "Nozzle Flow Characteristics in Pneumatic Force-Balance Circuits." *ASME Transactions,* (Feb. 1948).

nozzle back pressure. This opens supply port *H*, admitting air and thus increasing the output pressure. This pressure is fed back into the chamber *I* and forms the reference pressure of the nozzle system. It balances the increased signal pressure and drives the diaphragm *E* back nearly to its initial position. The increased pressure *B* also counteracts pressure *F* and restores the initial position of the diaphragm assembly.

Compression spring *L* adds to pressure *B* in balancing the diaphragm assembly against pressure *F*. It is therefore necessary that the nozzle back pressure is at all times above the output pressure by a fixed amount, which is equivalent to the spring force. This is the constant pressure differential across the nozzle.

This method assures that the increased input signal results in increases at *B*, *F*, and *I* with little displacement of original positions, and thus approaches the null system.

The schema of Figure 8–6 adds the mechanism required to also vary the supply pressure to the nozzle system with changes of the input sig-

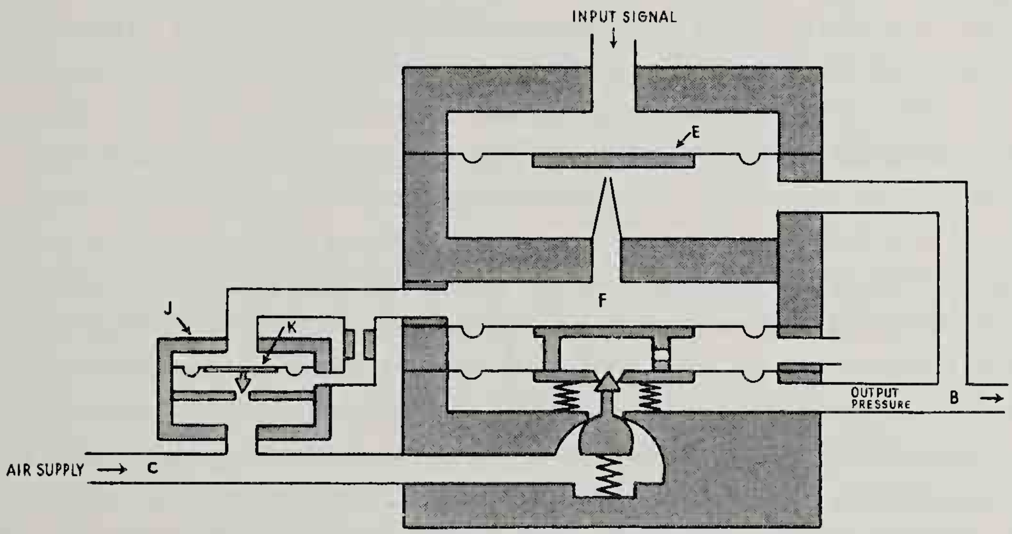


Figure 8–6. Pilot relay with variable supply and reference pressures.

nal. The air supply *C* does not pass directly to the fixed restriction, but flows first through the pressure regulator *J*. The diaphragm *K* controls the flow and pressure drop across the regulator. The position of the diaphragm is controlled by the nozzle back pressure *F*. As described above, the nozzle back pressure is proportional to the output pressure *B*. This makes p_s vary in proportion with p_t.

The fact that diaphragm *K* is now being deflected while the deflection of the *E* diaphragm is minimized should not be overlooked. The deflection is placed at a point where its effects are negligible.

The conclusions which can be drawn and are of rather general validity

are that (a) a null system can be designed for greater accuracy than a deflection system, (b) the deflection system is inherently a simpler arrangement than the null system, (c) the advantage of the null system may be approached by reducing the deflection. The foregoing also implies a generalized statement: Since a simpler system is always to be preferred, the deflection system should be chosen unless the greater accuracy obtainable with the null system is desired.

Position Balances

The choice between force balance and position balance is determined by the nature of input and output. If they both are forces, the force balance will be chosen. If they both are positions, the position balance will be preferred. If one is a force and the other a position, the decision becomes generally arbitrary.

The primary feedback from a controlled variable is practically always a force, either mechanical or electrical. Frequently, this signal is used not only for controlling, but for recording and indicating as well. In a recorder, for example, a pen displacement indicates the magnitude of the controlled variable. The output of a pneumatic recorder-controller is a pressure which applied to a certain area is a force. In the case of an electric controller it is generally an electromotive force (emf).

Ignoring for the moment the control-point-setting mechanism, the problem that arises is to make a choice among three possible arrangements. One is using the force of the primary feedback and comparing it with the force of the output signal. The other consists of using the positioning mechanism of the pen and comparing the position with the controller output signal by converting the latter from force into position. The third arrangement possible is to convert the displacement of the pen mechanism into a force and compare it with the force of the controller output signal. The first possibility is rarely, if ever, used in a recorder-controller. The main reason is that the force of the primary feedback depends on the nature and range of the controlled variable. It therefore does not allow a simple universal balancing mechanism. Furthermore, the force level is frequently extremely low and produces practical design difficulties. This makes the positioning mechanism the preferred input source for the balancing system.

In electric recorder-controllers, the third arrangement—converting the displacement of the positioning mechanism—is frequently preferred. For example, the displacement may be linked to the wiper of a potentiometer and thereby regulate an emf which is the controller output signal.

In pneumatic recorder-controllers the most widely used methods are those illustrated in Figure 8–7 where the controller output signal is con-

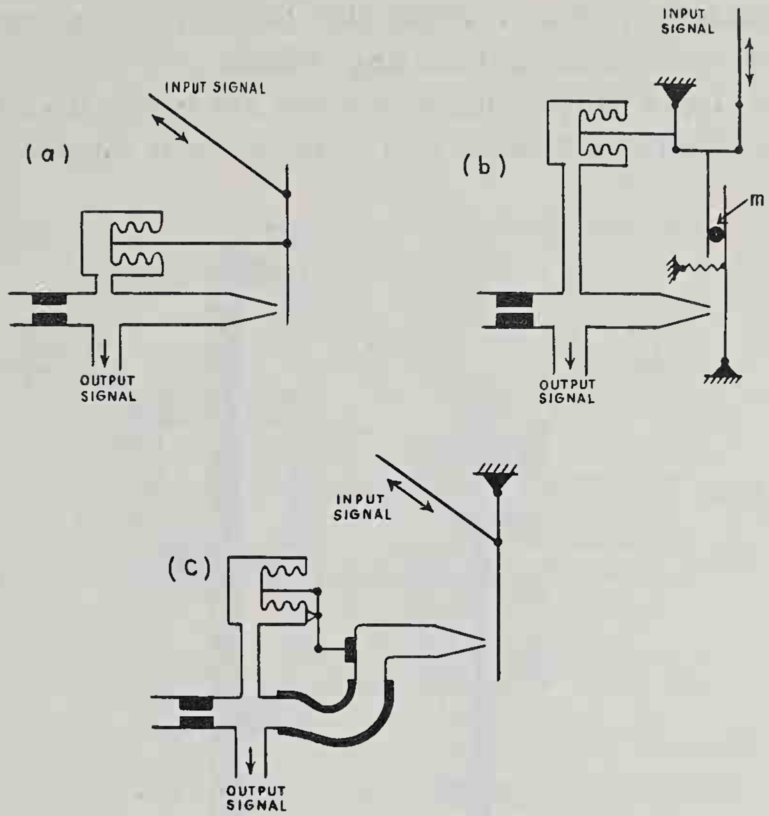


Figure 8–7. Three typical forms of pneumatic position balances.

verted into a bellows displacement; this is done with a position input signal derived from the measuring means which may simultaneously be used to position an indicator or pen.

Pneumatic Position Balances. Figure 8–7 illustrates three typical forms of pneumatic position balances. The flapper in case (a) is positioned by the input signal linkage which is connected with the pen or indicator positioning mechanism and by the linkage to the feedback bellows. This is identical with Figure 8–2.

Case (b) is a modification of (a) which allows, by means of the adjustable fulcrum *m*, the change of the proportional relation between input and output signal. The spring shown in the illustration has no balancing function but serves only to hold the flapper against the fulcrum.

Case (c) differs insofar as both flapper and nozzle positions are changed. The input signal positions the flapper and the feedback signal, the nozzle. The relative position of flapper and nozzle is the position balance which determines the pressure of the output signal.

Hydraulic Position Balances. The output signal of a hydraulic controller is essentially flow which is converted into change of position of an actuating piston. Hence a position balance is generally applied. Fig-

ure 8–8 illustrates a hydraulic controller, the main components of which are a four-way valve and an actuating cylinder.

Spool and sleeve of the four-way valve are both movable. Thus the spool may be displaced upward and if the sleeve is displaced upward by

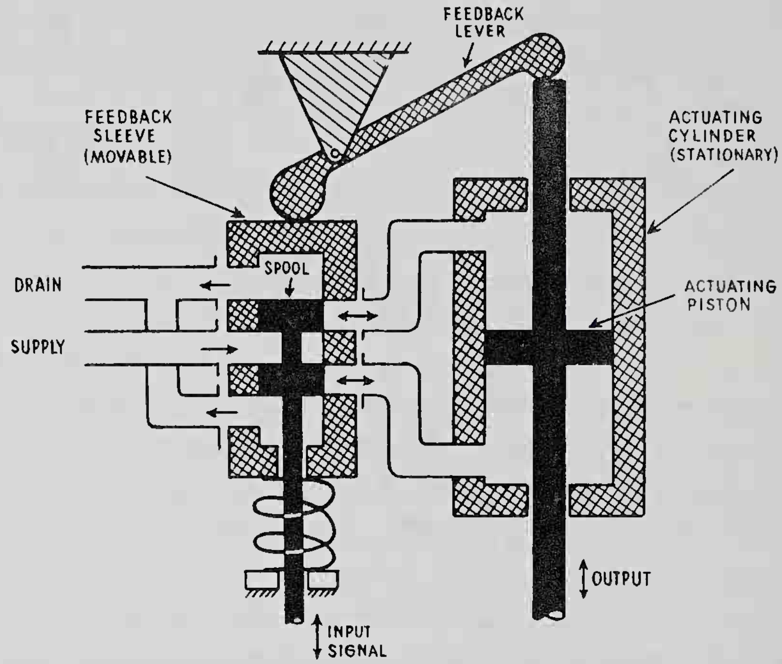


Figure 8–8. Hydraulic system position balance.

the same amount, their relative position is as before. The input signal displaces the spool. Suppose the change of the controlled variable is such that the spool moves upward. This will open the port to the top of the actuating piston connecting it with the oil supply. The port to the lower part of the actuating cylinder also opens, connecting from underneath the piston to the drain. As a consequence the actuating piston moves downward. The feedback lever holds on its extreme left the sleeve of the four-way valve against the force of a spring. As the actuating piston moves downward, the feedback lever follows the motion by rotating clockwise. This moves the feedback sleeve upward, restoring the initial relative position between spool and sleeve, i.e., closing the ports to the actuating cylinder. This stops the motion of the actuating piston.

The position balance between spool and sleeve thus assures accurate positioning of the actuating piston in proportion to the input signal.

Wheatstone Bridge

The most frequently used bridge circuit is the Wheatstone bridge. The schema of Figure 8–9 illustrates the principle.

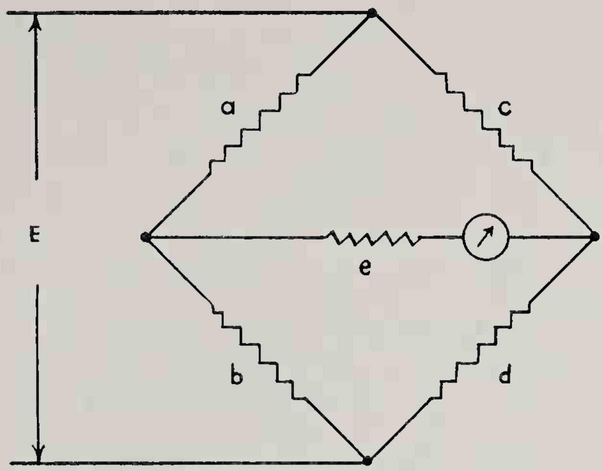


Figure 8–9. Wheatstone bridge.

The signal current I through resistor e is given by

$$I = E\frac{bc - ad}{ab(c + d) + (a + b)(ed + de + ce)} \tag{8–3}$$

The equation shows that for $bc = ad$ the current through resistor e is zero. If, for example, resistor a is a resistance temperature detector, the resistance of which changes with temperature, then the signal current can be measured by an ammeter with a scale calibrated in degrees Fahrenheit.

To obtain the maximum sensitivity of the bridge, all five resistors should be of approximately the same magnitude. If resistors c, d, and e are of value equal to b, then the equation (8–3) can be written

$$I = E\frac{b - a}{5ab + 3b^2}$$

The nonlinearity between change of resistance in resistor a and signal current flow is illustrated in Figure 8–10.

The nonlinearities of the Wheatstone bridge disappear when it is used as a null balance. The signal current that flows through resistor e is then the unbalance which is applied to a servo amplifier. A servomotor energized by the amplifier positions the sliding contact of a slidewire resistance which is one of the legs in the bridge. The null condition of $bc = ad$ is thus re-established after each disturbance of the bridge. If resistor a is the primary element which responds to changes in the controlled variable, then resistor c is readjusted by the servomotor when a changes.

The disadvantage of the described arrangement is that contact resistance of the sliding contact produces an erroneous signal. The method which prevents this is illustrated in Figure 8–11. The contact resistance

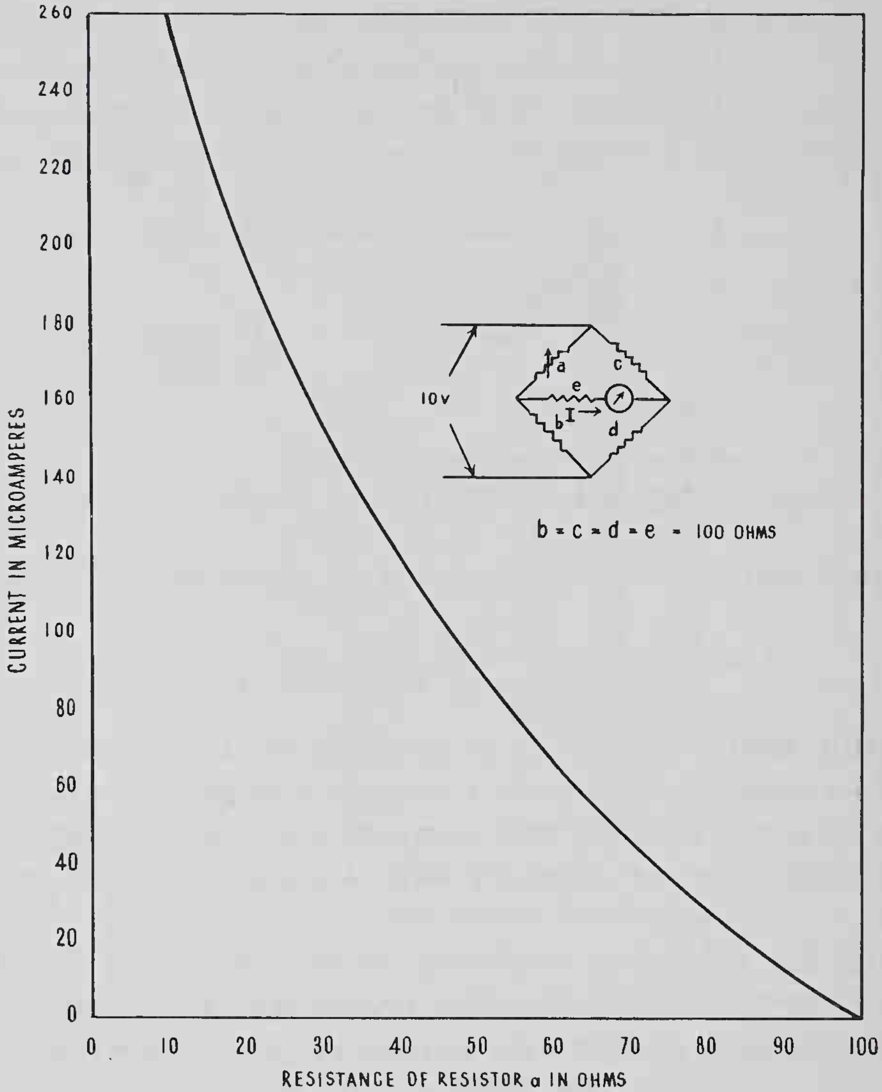


Figure 8–10. Current flow in Wheatstone bridge.

of slidewire *A* appears in the connection to the servo amplifier, and has no influence, since no current flows under balanced conditions.

The illustration shows a second slidewire. The positioning shafts of both slidewires *A* and *B* are coupled together and controlled by the same servomotor. The arrangement is used with resistance temperature detectors. Resistor *a* is the primary element which is exposed to the controlled temperature. It is located at a distance from the remaining bridge circuit. The connecting wires are therefore also exposed to atmospheric connections which may be subject to change and hence influence the wire resistance, producing an erroneous signal. In this case a three-wire connection is used, i.e., wires *m, p,* and *n,* are brought to the location of *a.* Resistance changes in wire *p* are again of no influence in a null-balanced bridge. Changes in *m* and *n* affect two different legs in the bridge and compensate each other. Since *a* changes with temperature, *b* must

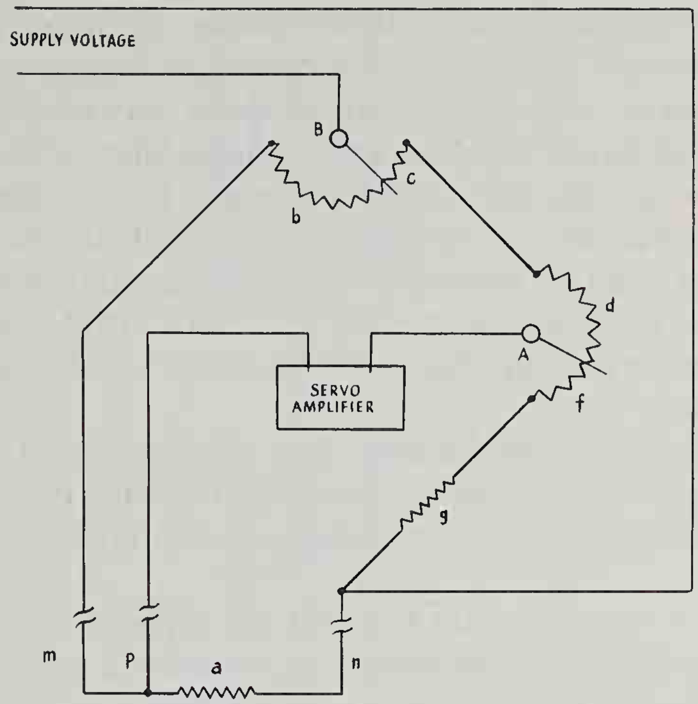


Figure 8–11. Self-balancing double-slidewire bridge.

be made to change correspondingly, otherwise the compensation is not complete. This change of b is obtained by means of slidewire B. The coupling between A and B provides that for any position

$$c + d = f + g$$

from which it follows that

$$\frac{f + g}{c + d} = 1 \tag{8–4}$$

For null condition in the bridge it is necessary that

$$\frac{a}{b} = \frac{f + g}{c + d}$$

Substituting equation (8–4) yields

$$\frac{a}{b} = 1 \text{ or } a = b$$

which is the condition for complete compensation.

In strain gauges, for example, temperature changes which affect the stress-sensitive resistor a must not be detected. In this case, a second resistor of the same material but not exposed to stress is mounted close to a. This resistor then represents a compensating leg in the bridge circuit.

Where slidewires are objectionable because of limited life expectancy other elements may be used. The Wheatstone bridge is not limited to resistance changes; capacitors or inductances may replace the resistors.

The null condition $bc = ad$ shows the adaptability of the null-balanced Wheatstone bridge for computing purposes. For example, let d be a fixed resistor where $d = 1$. In this case $a = bc$. If the servomotor positions resistor a, then its corresponding angular shaft position is in null position always equal to the product of b and c. Thus, resistances b and c may represent two controlled variables which require an output equal to their product.

Similarly, if $b = c$, which means both resistors are of the same magnitude and exposed to the same controlled variable, then $a = b^2$. In this case the servomotor shaft position is proportional to the square of the input.

Inversely, if $b = c$, and both resistors are adjusted by the servomotor, while a is exposed to the controlled variable, then $b^2 = a$, or $b = \sqrt{a}$. This means that the shaft position is proportional to the square root of the input.

Finally, if resistor b is fixed and made equal to unity, then $a = c/d$. If a is positioned by the servomotor then its shaft position is proportional to the ratio of inputs c and d.

The usage of bridge circuits for computation is a fine example of the possibilities inherent in a null balance. The Wheatstone bridge is used as position balance for many purposes, but for computation it is practically a prerequisite that it be a null balance.

Sorteberg Bridge. Equivalents to the Wheatstone bridge are used in pneumatic and hydraulic circuits. In fact, the push-pull flapper nozzle in Figure 6–12 is really a Wheatstone bridge.

A specific pneumatic bridge arrangement is the Sorteberg bridge, the principle of which is illustrated in Figure 8–12. Connections are shown for square root extraction. The input signal is applied at bellows *D*. An increase of the signal rotates lever *E* clockwise about roller *F*. This increases the back pressure of nozzle *G* and with it the output signal. This signal is also applied to bellows *A*, which results in an increase of the back pressure of nozzle *H*. The back pressure is applied to a Bellofram or equivalent, which pushes rod *M* downward and with it rollers *K* and *F*. This changes the fulcrums for levers *E* and *P*. The result is that both flappers tend to move away from their nozzles, counteracting the action of the input signal which tends to close the nozzles. A null balance is thus obtained.

The active forces are those due to the pneumatic pressures multiplied

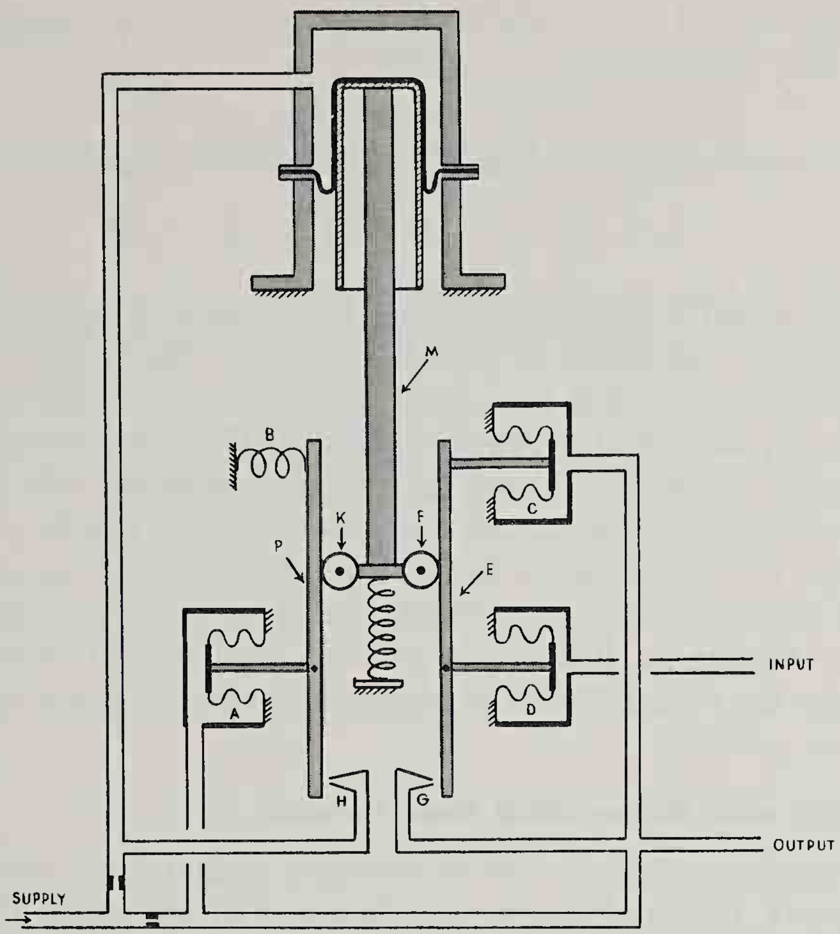


Figure 8–12. Schematic drawing of Sorteberg bridge.

by the bellows areas and by the length of the lever as determined by the position of the fulcrum. An additional force is produced by the spring. Since the deflection in the null balance is negligible, the spring force is practically a constant. Force balance is established when

$$A \times a = B \times b \text{ and } D \times d = C \times c$$

where a, b, c and d are the lever lengths associated with A, B, C, and D, respectively. Displacement of rollers K and F changes simultaneously the lengths of all four levers. However, at all times $b = c$ and $a = d$. Hence

$$A \times d = B \times c \text{ and } D \times d = C \times c$$

Dividing the first of these two equations through the second, yields

$$\frac{A}{D} = \frac{B}{C} \text{ or } A = D\frac{B}{C}$$

Furthermore, B is the constant spring force which may be assumed to

have unity. Since A and C are interconnected, they have the same pressure and $A = C$. Therefore,

$$A = D\frac{B}{C} \text{ may be written as}$$

$$A = D\frac{1}{A} \quad \text{or} \quad A^2 = D \text{ or } A = \sqrt{D}$$

The output signal is equal to A and the input signal is equal to D. Hence the output is proportional to the square root of the input.

It is quite obvious that the algebraic relationship is the same as for the previously described Wheatstone bridge. The same variations apply therefore. For example, multiplying of signals requires only that the air supply—via the restriction as shown—together with nozzle G, be cut off from A and C and connected to D. Applying one input signal to A and the other to C results in an output signal from D which is equal to the product of the inputs. If the same signal is applied at A and C, the output squares the signal. Likewise, connections for dividing and proportioning are possible.

Computing with Differential Transformers

Differential transformers can be arranged so that their output is equal to the square, the product or other function of the input. Figure 8–13

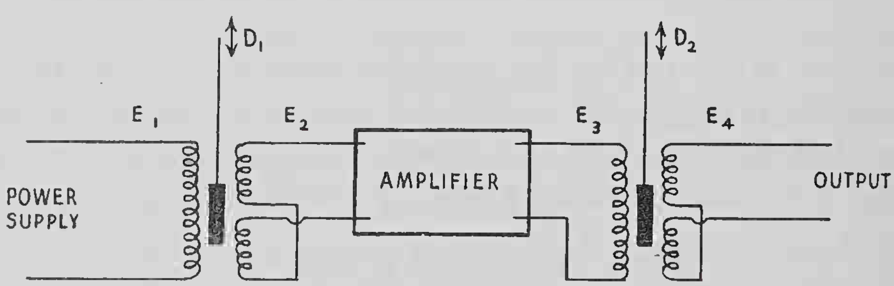


Figure 8–13. Multiplication with differential transformers.

illustrates the use of two differential transformers for multiplication. Voltage E_2 is proportional to E_1 and the displacement D_1 of the core. Thus

$$E_2 = k_1E_1D_1$$

The voltage E_2 is amplified to E_3, hence

$$E_3 = k_2E_2$$

Finally,
$$E_4 = k_3E_3D_2$$

By substitution

$$E_4 = k_1 k_2 k_3 E_1 D_1 D_2$$

or by combining the constants k_1, k_2, and k_3 into the single constant k,

$$E_4 = kE_1 D_1 D_2$$

which expresses the multiplication of the inputs D_1 and D_2. By making $D_1 = D_2$, the resulting output is the square of the input.

9. MEASURING ELEMENTS

Problems of measuring elements that have specific bearing on control are primarily square root extraction and temperature measurement. The square root extraction becomes important in flow measurement. Since in most cases flow is determined by measuring the pressure differential across an orifice plate, nozzle, or similar element, the magnitude is expressed as the square of the flow rate, since

$$Q = kA\sqrt{p_1 - p_2}$$

and $(p_1 - p_2)$ is measured. If the flow controller gain is expressed as change of output signal per unit change of flow, then the gain will change with the magnitude of the flow rate, because the flow rate is a nonlinear relation of the differential pressure. This change in controller gain is equivalent to changing the proportional band in a controller with proportional-position action. This has consequences in response speed and stability and is undesirable for optimum control although quite frequently it can be tolerated. More serious, however, is the nonlinearity in multiplications with other signals which are linear. This is necessary, for example, where gas flow is measured and corrections for changes in the gas pressure and temperature are required. In order to avoid all these difficulties, the square-root is extracted from the differential-pressure measurement.

Of particular importance in temperature measurement for automatic control is the time lag that occurs in the primary element. In practically all other cases the response time of the primary element is negligible in comparison to the response time in the process or occasionally, the final control element. In temperature control, however, the response time of the primary element has sufficient bearing on the dynamic response of the system to warrant separate discussion.

Square-Root Extraction

Three basic methods for square-root extraction are in general use: Ledoux bell, mechanical linkage, and electrical force balance.

Ledoux Bell. Figure 9–1 illustrates the principle of the Ledoux bell. The bell is floating, semi-immersed in mercury or some other suitable liquid. The mercury surface inside the bell is exposed to the high-pressure

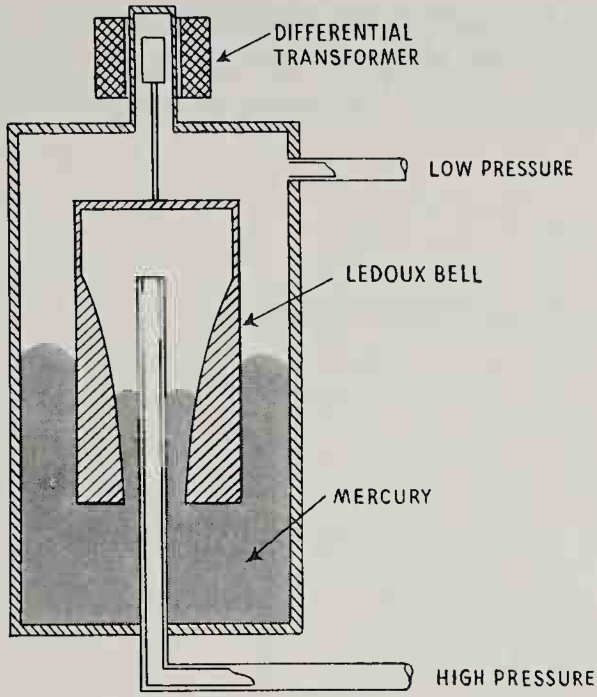


Figure 9–1. Principle of Ledoux bell.

side of a differential-pressure primary element. The surface outside the bell is exposed to the low pressure side. Changing the differential pressure will displace the bell correspondingly. The parabolic shape on the inside of the bell makes this displacement proportional to the square root of the differential pressure. The differential transformer thus picks up a signal which changes in linear relation with flow.

Limitations of this arrangement are the relatively high machining cost of such a bell and the fact that each bell is suitable for one range only. To change the measuring range, it is necessary to change bells. Furthermore, it requires mercury or some other suitable liquid and is not applicable to diaphragm and bellows type flow meters.

Mechanical Linkage. A cam can be shaped so that it converts a motion into practically any mathematical function. However, such cams must generally be made with high accuracy to render useful results. The relatively small motions which are usually available in flow meters result in additional difficulties. Furthermore, a cam introduces friction which is undesirable with the small forces available. However, cams are used for square-root extraction, particularly in ring-balance meters where motion and available forces are larger than in most other differential-pressure devices.

Figure 9–2 shows an arrangement in which the input force is balanced by a counterweight with the lever effect of the counterweight modified by a cam. The deflection of the beam is the result of two opposing

torques, one from the input force, the other from the weight. In the position shown in the upper diagram, the torque is proportional to L_1. In the lower diagram the input force has increased. The resulting torque

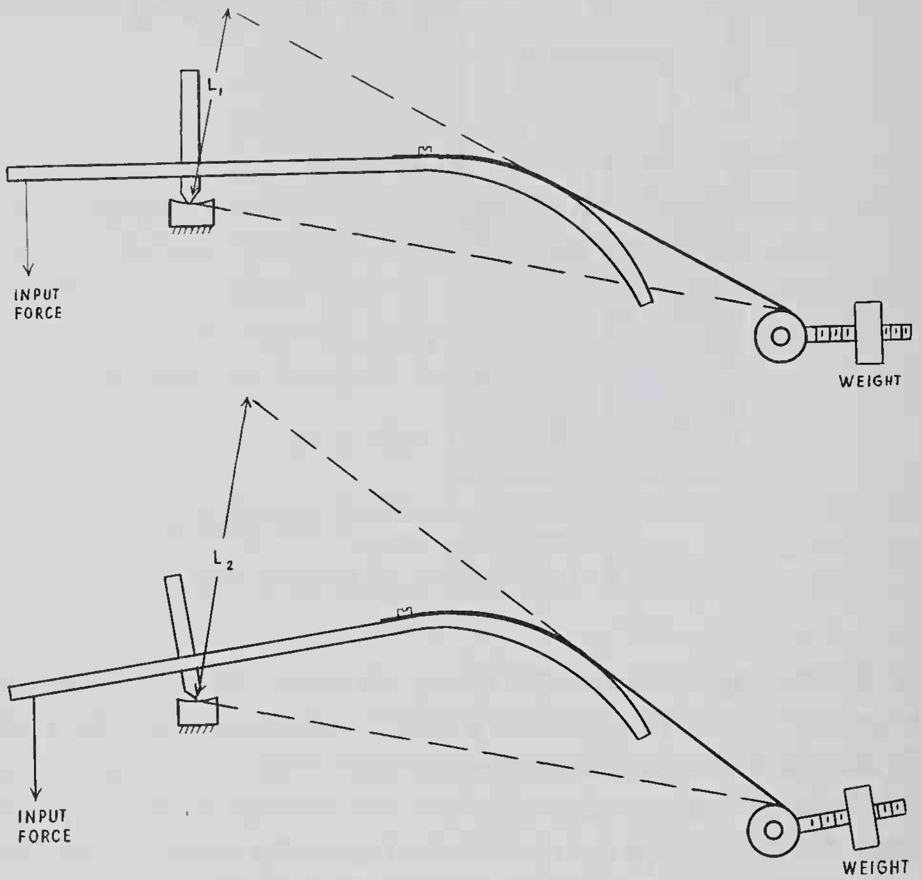


Figure 9–2. Cam and counterweight for square-root extraction.

rotates the lever counterclockwise. This increases the effect of the counterweight in proportion to $L_2 - L_1$. A new balance is obtained with a deflection as shown. The cam can be shaped to obtain a deflection proportional to the square root of the input force.

Figure 9–3 shows a lever system for square-root extraction. Levers *A* and *C* both follow circular paths as outlined. The input and output signals are assumed to be angular displacement. If the input is a shaft at *E,* the rotation of which moves lever *C* and hence levers *B* and *A,* then a second shaft at *D,* connected to lever *A,* will provide an output signal which can be made a square-root function of *E,* provided that certain dimensions and angular displacements are observed.

The angles through which the input shaft *D* and the output shaft *E* move, are called α and β, respectively. The purpose of the arrangement in Figure 9–3 is a relationship in which

$$\alpha = k \sqrt{\beta}$$

where k is a proportionality constant. This relation holds true in Figure 9–3 for a maximum angle of 100° for either shaft. When shaft D moves through the maximum angle $\alpha = 192.5 - 92.5 = 100$, shaft E moves

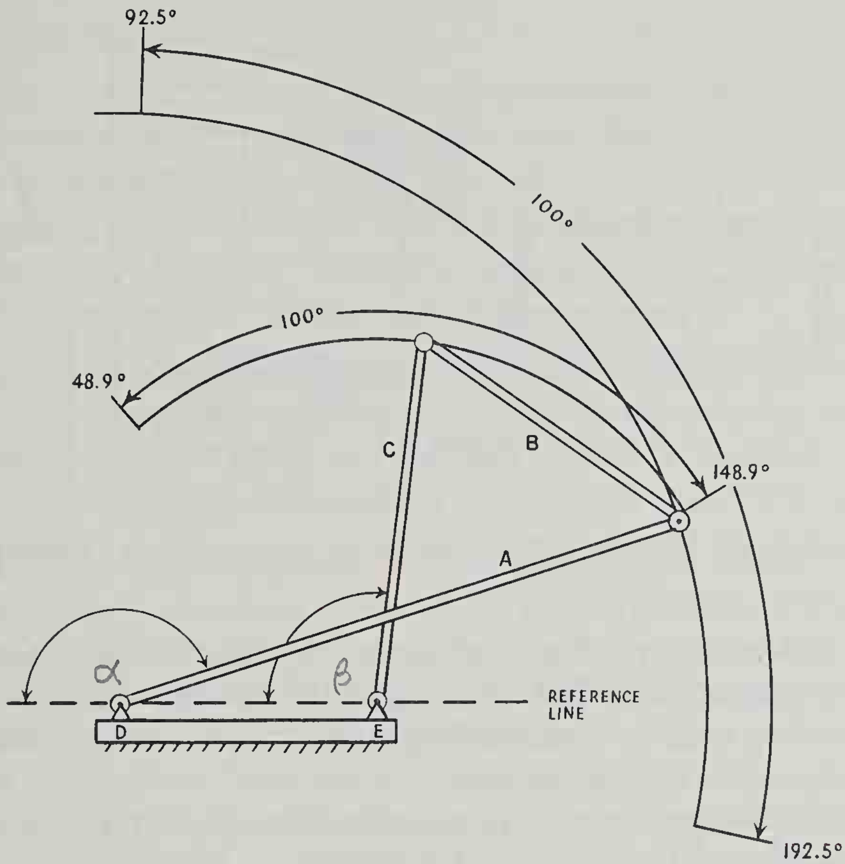


Figure 9–3. Lever system for square-root extraction.

through $\beta = 148.9 - 48.9 = 100$ degrees. Since $100 = k\ \sqrt{100}$, it follows that $k = 10$.

In the position shown in the illustration, $\alpha = 163.2 - 92.5 = 70.7$ and $\beta = 98.9 - 48.9 = 50$. This corresponds with the desired relationship because $70.7 = 10\ \sqrt{50}$.

Certain dimensions must be maintained in the layout of such a mechanism in order to obtain a square-root relation. These dimensions are as follows:

Distance between D and E	1 in.
Length of lever A	2.291 in.
Length of lever B	1.227 in.
Length of lever C	1.409 in.

Any multiple of these dimensions may, of course, be used.

The disadvantage of this and similar linkage mechanisms is that high

precision is required in manufacturing them in order to obtain reasonable accuracy.

Electrical Force Balance. The method of obtaining square-root extraction by electrical force balance is illustrated in Figure 9–4. The force

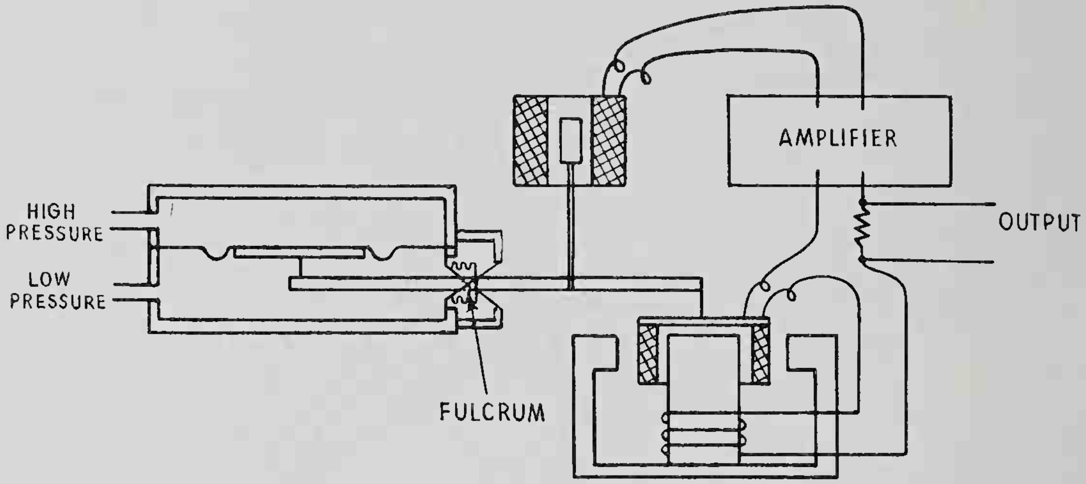


Figure 9–4. Square-root extraction by electrical force balance.

due to the differential pressure across the diaphragm is balanced by the force in the moving coil. The winding of the electromagnet is in series with the moving coil, so that the force that the coil exerts on the beam is equal to the square of the current.

The differential pressure is proportional to the square of the flow, hence the corresponding force on the beam is

$$F_1 = k_1Q^2$$

The moving coil force is

$$F_2 = k_2i^2$$

Since $F_1 = F_2$, and the square root of the proportional constants k_1 and k_2 can be expressed by k, it follows that

$$i = kQ$$

which means that the current flowing through the moving coil is proportional to flow, i.e., the square root of the differential pressure.

The differential transformer picks up any motion in the beam. Its output is connected to an electronic amplifier and a corresponding signal from the amplifier is applied to the moving coil system. The combined gain of the differential transformer and amplifier is so large that the beam motion required to increase the signal from minimum to maximum can be made negligible for all practical purposes. This has the further

advantage that it minimizes those nonlinearities which increase with diaphragm deflection.

The output is taken as a voltage across a resistor in the feedback from the amplifier to the moving coil system.

Temperature Control

Of the three most frequently controlled variables—temperature, pressure, flow—temperature is generally the most difficult to control.

Inaccuracies due to structural and physical limitations of the measuring means are disregarded here. The inherent difficulties of temperature measurement are however discussed. Two groups of errors can be distinguished in temperature measurement:

1. Static errors due to inaccuracies of reproduction of the temperature under measurement by the temperature of the primary element;
2. Dynamic errors due to thermal lags caused by thermal resistances and capacitances in the primary element, its protective wells, etc.

Temperature reproduction. The temperature to which the measuring means actually responds is that of the primary element. Only through the primary element—hence indirectly—is the temperature of the controlled medium measured. Static errors in the control system result from lack of accuracy in temperature reproduction.

Heat is transmitted from the surroundings to the primary element by conduction, convection, and radiation. A room thermostat, for example, is generally supposed to respond to changes in room temperature brought to it by heat conduction. The fact is that exposure to draft, i.e., convection, will result in faulty action of the thermostat. The reason is that the thermostat is intended to respond to conduction alone, but can obviously not distinguish between conduction and convection. Similarly, in the measurement of hot gases, e.g., in furnaces, or in the uptakes and regenerators of open hearth steel furnaces or glass tanks, the thermocouple or—more commonly—its protective well, will pick up radiant heat from hot surroundings (walls, etc.) and thus produce erroneous readings.

Primary elements are available that reduce the effects of undesirable radiation to a minimum. Frequently, however, simple shielding of the primary element suffices to eliminate much of the error that otherwise results.

Another static error is caused by the loss of heat that flows along the primary element into the surrounding atmosphere. Figure 9–5 shows the installation of a well in a pipeline. In cases where the temperature difference between the controlled medium and the atmosphere is large,

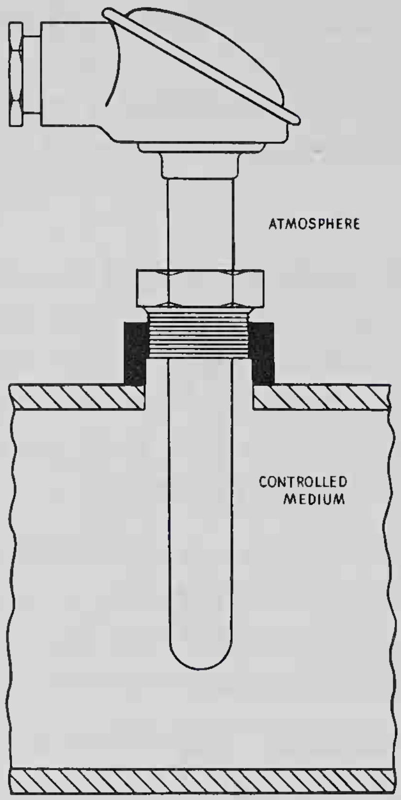


Figure 9–5. Installation of a well in a pipeline.

heat flow along the well may actually lead to a reduction of temperature even in those parts of the well that are immersed in the controlled medium. This may lead to a temperature indication by the primary element which is too low. In cases of gas and vapor measurements with considerable temperature differences between atmosphere and controlled medium the following precautions should be taken:

1. Lagging of the well exposed to the atmosphere;
2. Deep immersion of the primary element into the controlled medium. The immersion depth should be at least 1½ to 2 times the length of the sensitive part of the element when the fluid velocity is high, and twice that much when it is low. Another rule is 6 to 8 times the well (or protection tube) diameter for high and 12 to 16 times for low velocities.
3. Heating of the well exposed to the atmosphere, possibly by branching off some of the process fluid.
4. If oil, mercury, or similar material for better heat transmission is used, as described further below, such filling should not go beyond the temperature-sensing element, since it otherwise contributes to heat conduction away from the element.

It happens occasionally in very fast flowing gases that temperature

signals of the primary element are on the high side. This is mainly due to friction. Air flowing at 10,000 feet per minute may result in an error of about 2°F, which rises however to 40°F at 40,000 feet per minute.

The error of temperature-measuring elements due to their location in air pockets, under conditions of impinging flames, etc., is well known. The conclusion regarding temperature reproduction by the primary element is that conditions and location of the installation must be considered in order to obtain a correct primary feedback signal.

Thermal lags. Thermal lags affect the dynamic response of the control system. Consider a filled-system thermometer bulb as shown schematically in Figure 9–6. Suppose the bulb has been moved from a cold

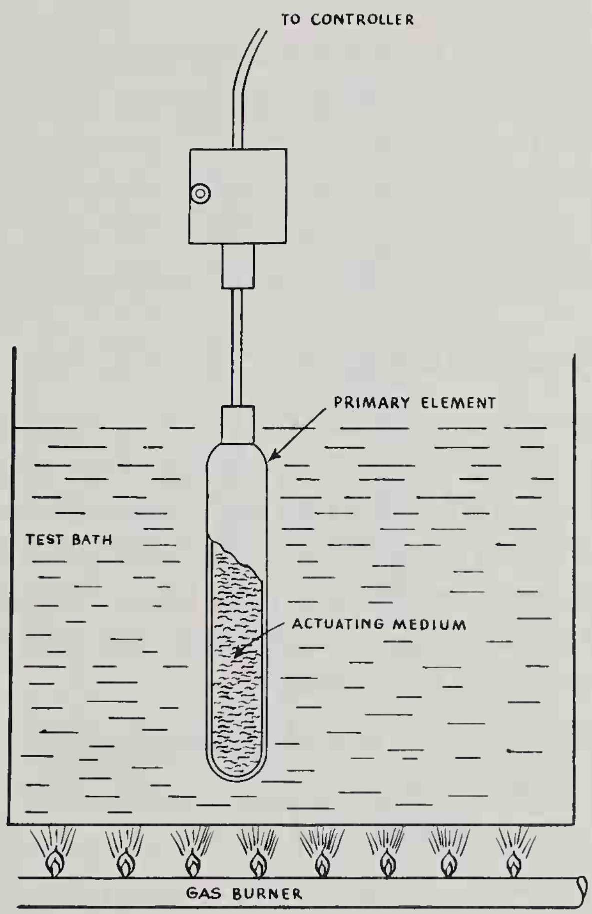


Figure 9–6. Filled-system thermometer bulb.

medium into the tank bath as shown, which is at an elevated temperature. This corresponds to a step change. Heat flows then through the walls of the bulb to the actuating medium until the temperature of the actuating medium is equal to that of the bath. Only then will the measuring means settle at the new temperature.

The following discussion investigates the rate and quantity of heat flow that takes place until the thermal balance is reestablished.

Heat *quantity* is generally expressed in British thermal units—Btu—defined as the heat required to raise the temperature of one pound of water by one degree Fahrenheit. Heat *capacity* expresses how much heat in Btu has to be transmitted to one pound of a substance in order to raise its temperature by one degree Fahrenheit.*

The amount of heat required to raise the heat of the primary element to that of the bath depends on the heat capacity of the primary element. It is given by the equation

$$Q = c_p W(T_a - T_b) \tag{9-1}$$

where

Q is the amount of heat in Btu to be stored in the primary element
c_p is the heat capacity of the primary element in Btu/(lb)(°F)
W is the weight of the primary element in pounds. Since the primary element contains several substances of different weights and heat capacities, it is necessary to establish the $c_p W$ product for each substance and then add the various products to obtain a single $c_p W$.
$(T_a - T_b)$ is the temperature rise from T_b *to* T_a.

The heat flow into the capacities of the primary element in Figure 9–6 has to overcome a certain thermal resistance. This resistance exists not only in the well of the bulb—in fact this is the smallest contribution—but mainly in the film on both sides of the bulb's metallic surface.

A fluid is an extremely poor heat conductor. For example, at room temperature, the following figures illustrate the poor conductance of water and air as compared with steel:

Steel	26	Btu/(hr) (sq ft) (°F/ft)
Water	0.36	
Air	0.016	

Steel conducts over 70 times as much heat as water and over 1600 times as much as air.

However, fluids produce convection currents in partial compensation for their poor conductance. This is due to the formation of layers of dif-

* The term specific heat is frequently used in the same connection. Specific heat compares the heat capacity of a given substance to that of water and is expressed as a ratio between the two. Since, however, the heat capacity of water is approximately unity, it follows that numerically specific heat and heat capacity are identical for all practical purposes.

ferent temperatures. If the warmer layers are in the lower regions then they rise because of their lesser density. In rising they propagate heat much faster than would be possible by conduction only. This kind of heat propagation is referred to as *natural convection,* or simply convection.

Convection currents, however, slow down considerably as the fluid gets nearer a wall because of the friction between the wall surface and the fluid. Close to the surface there are no longer convection currents, and a sluggish film clings to the surface, through which the heat has to pass by conduction. In spite of the relative thinness of the film, its resistance to heat flow is far greater than the resistance of the bulb wall.

Figure 9–7 shows a cross-section through films and bulb wall of a primary element similar to Figure 9–6. Suppose it were possible to measure

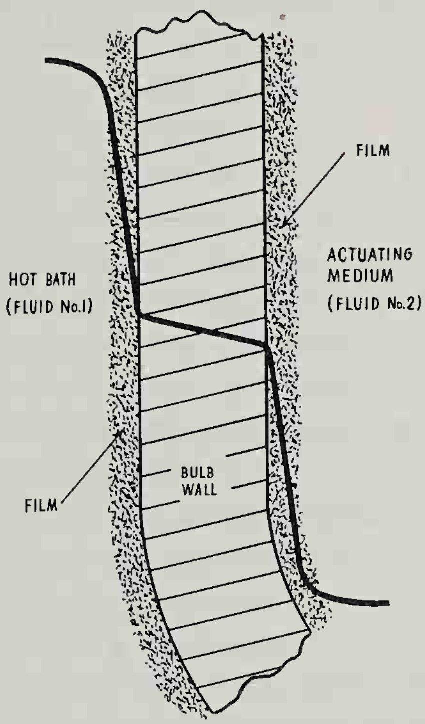


Figure 9–7. Cross-section through a piece of bulb wall with films and temperature gradient.

the temperature at various points in the fluids, films, and bulb wall, and that these measurements could be done shortly after the bulb has been changed from a cold to a hot bath. The various temperatures thus measured could then be represented by a graph showing how the temperature changes through the various points of the cross-section. Such a temperature gradient is illustrated in Figure 9–7, showing the sharp temperature drop through the film, expressing its high resistance, and the relatively small drop in the bulb wall itself.

The resistance of the film decreases with the velocity of the convection, the density of the fluid, the thermal conductivity of the fluid, and to a lesser degree with the heat capacity of the fluid. It increases slightly with the diameter of the bulb, and the viscosity of the fluid.

Natural convection is not the only form of convection. By means of stirrers, etc., a fluid may be agitated and thus the speed of heat propagation can be increased by so-called forced convection. The possibility of increasing the rate of heat flow by locating the primary element where convection currents are greatest (excepting those extremes where the very speed of those currents contributes to the error, as previously described), or by introducing forced convection should be given first consideration in any attempt to reduce the thermal lag in temperature control systems.

The lag in heat transfer increases with the protection required for the primary element. Thermocouples, for example, generally require wells to protect the thermocouple and to allow its removal in service without interrupting the process. In this case, the air pocket in the well, and the poor convection within it, slows down the response.

Hornfeck* compared responses of two thermocouples. Both were of No. 20 wire, iron-constantan couples, in brass wells of ⅝ in. outside and 7/16 in. inside diameter. In one case the thermocouple wires were not touching the wall, in the other they were silver-soldered to it. He found that measuring the temperature of air at 1150°F would take less than one half the time with the thermocouple silver-soldered to the wall and in agitated water it would take less than 4 per cent of the time it would take with the wire not touching the wall. This illustrates the effect of the air pocket. Of course, with a thermocouple silver-soldered to the wall of the well, one of the purposes of the well—to make the thermocouple easy to remove—is lost. Some commercially available arrangements to improve the heat transfer will be discussed farther below.

Occasionally two protecting tubes instead of one are required. This adds considerably to the temperature lag. Similarly, in pressure-actuated bulbs, a well is used over the bulb for extra protection and easy removal of the bulb. In such cases, the interspace should be filled with oil, mercury, or a similar liquid. Improved thermal contact may also be obtained by inserting metallic foils in the air space.

The equation that expresses heat flow through a thermal resistance is

$$q = hA\,(T_a - T_x) \tag{9–2}$$

---

* Hornfeck, A. J., "Response Characteristics of Thermometer Elements," *ASME Transactions*, (Feb. 1949).

where

q is the rate of heat flow in Btu/hr,
h is the heat-transfer coefficient, comprising films, wall and other paths of thermal conductance, in Btu/(hr) (sq ft) (°F),
A is the cross-sectional area through which the heat flows,
$T_a - T_x$ is the temperature difference of the two endpoints between which the heat flows.

The product hA is the total heat conductance between the controlled variable and the primary element.

This equation shows that to increase the rate of heat flow for a given temperature difference either the area or the heat-transfer coefficient should be increased.

The heat-transfer coefficient is largely controlled by the film, as has been discussed. Changing the bulb material or reducing its wall diameter may therefore be of benefit for the heat capacity, but has little influence on the overall heat-transfer coefficient.

Another recourse is, however, to use a bulb of large area. This is frequently done in gas temperature control and particularly where convection currents are slow. Various types of large-area bulbs are available.

Time Constant and Dead Time. As has been defined before, the time constant is the time required to produce 63.2 per cent of the total change. This is a convenient figure by which to express the dynamic action of a temperature-measuring primary element and which allows comparing the performance of similar elements.

In connection with temperature measurement, the speed of response is frequently expressed in response time instead of the time constant, and is defined, for example, by the *Scientific Apparatus Makers Association* (SAMA Tentative Standard RC3–12–1955) as follows: "The SAMA response time is the time required for the instrument to achieve 90 per cent of the change which the instrument is going to make as a result of an abrupt change in the measured quantity." Conversion from the time constant to the 90 per cent-response time is readily made by multiplying the time constant by 2.3. In the following discussion, the time constant will be used exclusively.

A large number of measurements have been made by various researchers. Unfortunately, their results are widely divergent. Figures given in the following discussion are an attempt to correlate these measurements, but while they may be valid for certain cases they can at best be considered general behavior patterns and are not meant to be specific data for universal application.

The smaller the time constant, T_1, the faster can the controller be made to respond to process deviations. The time constant is given by

$$T_1 = 60\frac{c_p W}{hA} \tag{9-3}$$

The coefficient 60 is required to convert time units, if the time constant is to be expressed in minutes.

Tests to determine time constants are not necessarily based on step changes. Occasionally, the response of a primary temperature element is shown graphically as illustrated in Figure 9–8 where the temperature of the test bath is gradually raised. The change of the bath temperature is expressed by graph a and the response of the primary element by graph b. The time constant can be obtained from such graphs by means of the equation

$$T_1 = E/m \tag{9-4}$$

where E is the dynamic error in °F and m is the rate of change of the temperature in the test bath in °F/min. For example, the dynamic error in Figure 9–8 is 15°F and the rate at which the test bath temperature

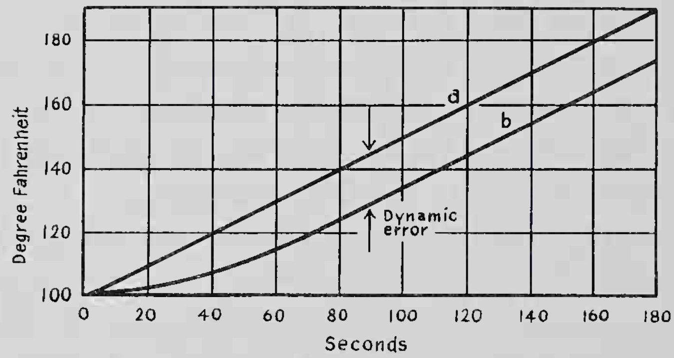


Figure 9–8. Response of primary element to gradual increase of test bath temperature.

changes is 30°F/min. Consequently, the time constant is 15/30 = 0.5 min.

This form of presentation illustrates the effect of thermal lags particularly well, since the dynamic error expresses the quantity by which the controller lags in its response to temperature changes, and hence may contribute to instability in the control system. It should be noticed that according to equation (9–4), the dynamic error increases with the rate of change of the test bath temperature which in a control system corresponds to the controlled variable.

It follows that the dynamic error is proportional to the rate of change of the controlled variable. If the controlled variable changes slowly enough, then the thermal lag will be of no adverse consequence. It may mean unnecessary expense to attempt reduction of thermal lags in a control system for which only slow temperature changes are to be expected.

Similarly, if the time constant of the process is considerably larger than what can be expected even of a slow primary element, no benefit can be derived from selecting a high-speed element. This is in general typical for fractionating columns where the temperature control system may have to cope with time constants of 40 to 60 min. Another example are annealing furnaces with time constants of 10 to 20 min. For heating systems, the time constant is usually between 1 and 5 min.

The well, interspace, bulb, etc. are all capacitances. It has been pointed out before that with multi-capacitance systems there results the assumption of dead time in addition to the time constant. For primary temperature elements in wells this dead time amounts to about 6 to 7 per cent of the time constant for water and to about 2 to 3 per cent for air.

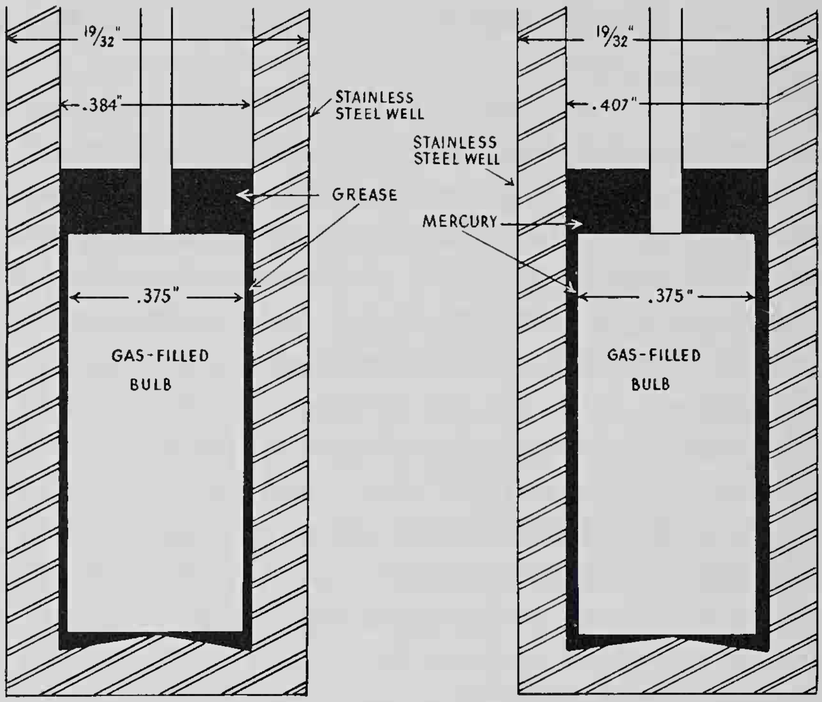
In view of rather considerable dead time in most temperature-controlled processes, the dead time of the primary element can usually be neglected. Typical dead times in temperature-controlled processes are as follows:

Process	Dead Time in Min.
Ammonia absorbers	8–10
Annealing furnaces	1–3
Fractionating columns	1–10
Heating systems	1–5
Oil tube stills	3
Superheaters	2

Linahan* measured the time constants of gas-filled bulbs using two configurations as illustrated in Figures 9–9*a* and *b*. The two gas-filled bulbs are identical. The inside diameter of the well, however, is smaller in case (*a*) than in (*b*). In one case, the interspace is filled with grease, in the other, with mercury. Time constants were very nearly the same so that for practical purposes, they can be considered identical. They

* Linahan, T. C., "The Dynamic Response of Industrial Thermometers in Wells," *ASME Transactions,* (May 1956).

were approximately 4 seconds for water with forced convection and 130 seconds for air flowing at 1400 ft/min., which corresponds with normal air duct velocities.



(a) Inside diameter of well 0.384 inch
Interspace filled with grease

(b) Inside diameter of well 0.407 inch
Interspace filled with mercury

Figure 9–9. Dimensions of two bulbs used in Linahan's measurements.

For arrangements which are similar to the one used in Figure 9–5, the following time constants may be used as guide:

Controlled Medium	Time Constant in Min.
Water and dilute solutions with forced convection	0.04–0.12
Saturated high-pressure steam, molten metals, water, and dilute solutions with natural convection	0.12–0.5
Oils and saturated steam at atmospheric pressure	0.3–1.2
Air at normal duct velocities	1–3
Heavy oils, syrups, etc.	2–7
Air with natural convection	6–30

The difference in time constants for the various types of filled-system thermometers, i.e., gas, vapor, liquid, and mercury, are small enough to be ignored. Measurements taken by the Brown Instrument Company* showed an average time constant of 0.1 min. for bulbs without well immersed in water at natural convection.

Figure 9–10 shows an interesting result from the same series of measurements. Curve (a) shows the response of a vapor-filled system

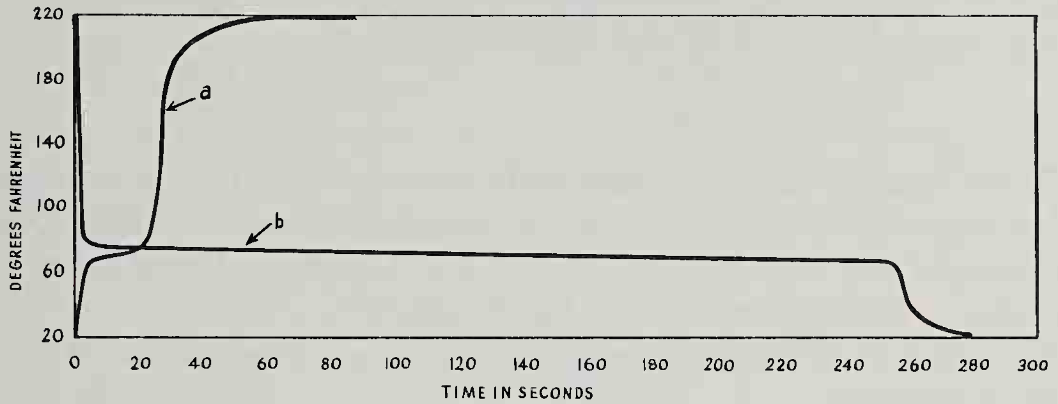


Figure 9–10. Response of a vapor-filled system at cross-ambient conditions.

which is transferred from a temperature lower to one higher than the temperature to which the connecting tubing and the measuring spiral are exposed. Curve (b) shows the reverse case, the response after a transfer from hot to cold. Obviously the temperature around connecting tubing and measuring spiral is about 65°F. As the actuating medium in the bulb reaches this temperature, the fluid outside the bulb starts to liquefy. The transition from the vapor to the liquid phase, and vice versa, takes time as expressed in the illustrated time delay. It will be noticed that this time delay is much longer when the temperature is lowered than when it is raised. The explanation is that with the rising temperature only the interface between liquid and vapor has to be brought to the new temperature. With a decreasing temperature all the vapor in the bulb must first be lowered to the new temperature.

This illustrates that control under such conditions would be unreliable. The ordinary Class II (vapor-filled) thermometers should never be used under cross-ambient conditions. There are, however, dual-filled Class II systems available which circumvent this shortcoming. In dual-filled bulbs, the volatile liquid and its vapor cannot leave the bulb. They only

---

* "Response Speeds of Pressure Type Thermometers," Bulletin No. 60–1, published by The Brown Instrument Company, Philadelphia, Pa., (1942).

partially fill the bulb space. The remaining space, as well as connecting tubing and measuring spiral, are filled with a non-volatile liquid. The fluids do not intermix. These bulbs are well suited for any temperatures within their range.

Hornfeck* made a number of measurements on resistance thermometers and thermocouples. Using them in air at approximately 1000 ft/min. gives the following time constants:

Resistance thermometer	0.07 min.
Iron-Constantan thermocouple	
No. 14 wire	0.13 min.
No. 19 wire	0.05 min.

These time constants are sufficiently close to consider the response of either resistance themometers or thermocouples as equally fast. It is interesting to see that the thinner wire, i.e., less mass, hence lower heat capacity, results in considerably faster response.

The effect of different capacities is even more pronounced in a comparison between

(a) a resistance thermometer in a stainless steel well of 0.600 outside and 0.563 inside diameter

with (b) a thermocouple of No. 12 wire in a stainless steel well of 0.563 outside and 0.438 inside diameter.

The time constants in min. are approximately as follows:

	Thermocouple	Resistance Thermometer
Air at 1000 ft/min.	2.27	0.67
Air at 5300 ft/min.	1.20	0.33
Water at 60 ft/min.	0.95	0.12

The mass of the thermocouple well is about three times that of the resistance thermometer well. Besides this, the comparatively heavy thermocouple wire adds even more to the larger capacitance. This illustrates that a comparison between two systems can be made only if they are operated under identical conditions. The contribution of the capacitance can also be seen from the above figures because the difference in time constant is largest where the capacitance/resistance relation is also largest, i.e., in water.

* *l.c.*

This is even more noticeable by increasing the air velocity from 1000 to 5300 ft/min. Since the resistance decreases with the higher velocity, i.e., the capacitance/resistance ratio increases, the thermocouple time constant increases from about 3.40 to 3.65 times the time constant of the resistance thermometer.

The decrease of thermal resistance due to increased speed of convection currents in gases has been discussed before. This is not limited to convection currents. Thus, when measuring air temperature, the thermal resistance and hence the time constant depends on the velocity of air that passes over the measuring element. This is illustrated in Figure 9–11. For

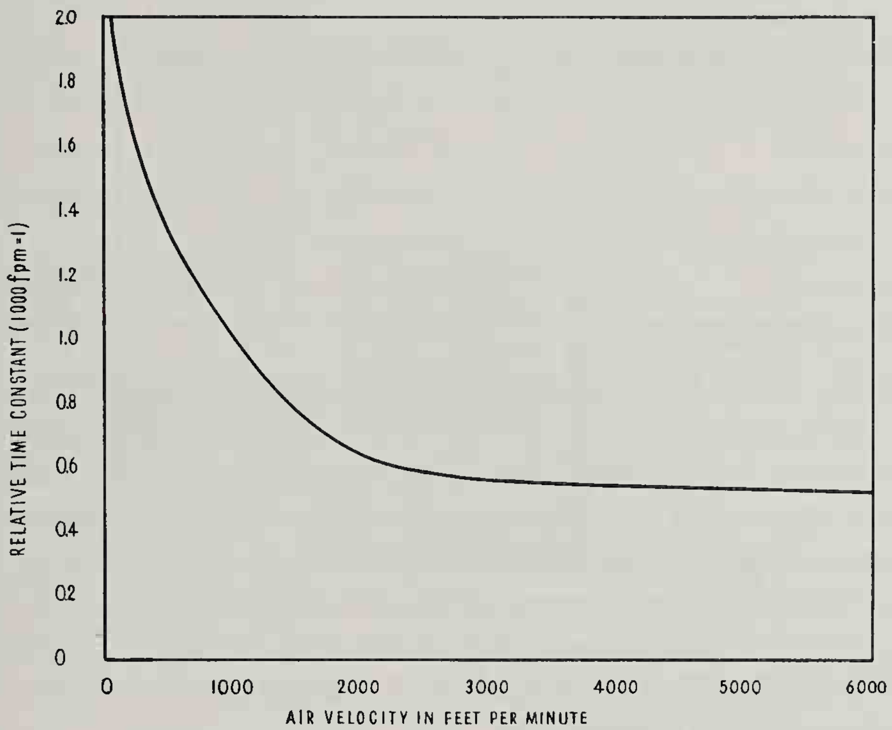


Figure 9–11. Change of time constant with air velocity.

example, the relative time constant for 600 ft/min. on this diagram is about 1.1, and for 5300 ft/min. about 0.5. Hence, if an actual time constant of 3 minutes has been determined for 600 ft/min., then the corresponding time constant for 5300 ft/min. would be $(0.5/1.1) \times 3 = 1.36$ min.

An attempt has been made to establish a table for the time constants of thermocouples and resistance thermometers. In view of the various factors that have been shown to control the time constant, the limits through which each time constant may vary must be consequently large. Correlation of available data gives the following approximations:

Controlled medium	Time Constant in Min.		
	bare	with metal protection tube	with ceramic protection tube
Water and dilute solution with forced convection	0.0005–0.05	0.01–1.2	0.2–3.0
Saturated high-pressure steam, molten metals, water and dilute solutions with natural convection	0.01–0.1	0.1–5.0	0.3–5.0
Oil and saturated steam at atmospheric pressure	0.02–0.2	0.2–7.0	0.7–10.0
Air at normal duct velocities	0.03–0.3	0.3–10.0	2.0–12.0
Heavy oils, syrups, etc.	0.08–0.8	1.0–20.0	4.0–20.0
Air with natural convection	0.15–1.5	2.0–60.0	8.0–60.0

Comparing this table with the values given previously for filled-system thermometers, it will be noticed that for thermocouples and resistance thermometers in metal protection tubes, time constants are both shorter and longer than with filled-system thermometers. The explanation is that a thermocouple or resistance thermometer hanging free in a well is slower in response than the usual filled-system thermometer, but that high-speed elements have been developed that have shortened the time constants considerably. Nevertheless, filled-system thermometers have been made faster in response too.

Three basic approaches are available to design a temperature-measuring primary element with shorter time constants:

1. increase of outside surface area
2. decrease of mass
3. improvement of thermal contact within the element.

A few examples will now be given where one or the other of these approaches has been used, first in filled-system thermometers, then in thermocouples, and last in resistance thermometers.

Figure 9–12 shows a capillary coil bulb. The extended length of such a bulb is about 30 ft, which obviously results in a large increase of area over an equivalent 6 in. standard bulb. The disadvantage is that it cannot be used with a well without losing its advantage of large outside

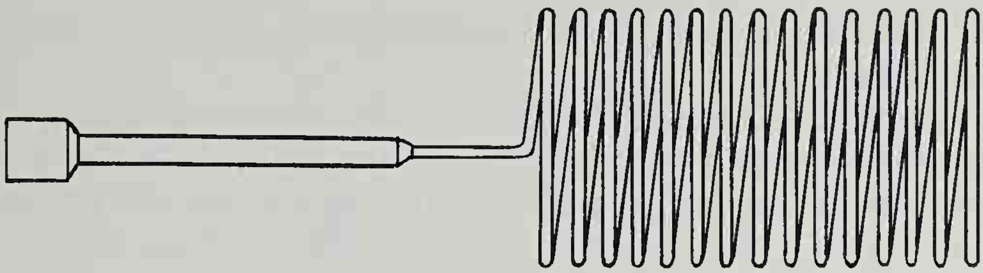


Figure 9–12. Capillary coil bulb.

surface area. Hence, it lacks mechanical protection and the ease of removal of a standard element. However, in air ducts, dryers, etc. it may be used to great advantage.

The Taylor Transaire temperature transmitter is illustrated in Figure 9–13. This design permits using a considerably smaller primary element

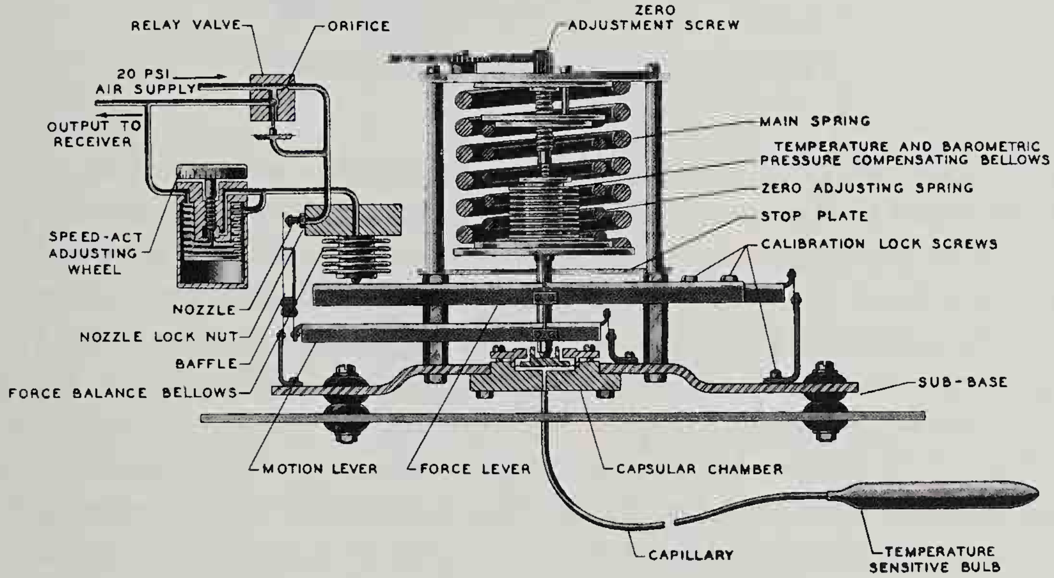


Figure 9–13. Transaire temperature transmitter.
(Courtesy of Taylor Instrument Cos.)

than others. The actuating medium in a filled-system thermometer is not only inside the bulb, but also outside in the connecting tubing and measuring spiral. This means that not only the bulb but other parts as well are temperature sensitive. This can be minimized by making the ratio of actuating medium inside bulb to actuating medium in connecting tubing and measuring spiral as large as possible. This results in a certain minimum bulb volume. By placing, however, a transmitter right at the primary element and converting the element's response into an air pressure signal, the volume outside the bulb becomes much reduced

and the bulb can be made correspondingly small. The bulb sizes are as small as ⅜ by 3 in.

Furthermore, this transmitter is built to provide rate action. Due to the relatively long time constants, rate action in the controller is generally desirable. In this case, it is built into the transmitter and the primary feedback signal has the corresponding correction. Rate action in the controller is not required.

Figure 9–14 shows a pencil type thermocouple which makes the iron protection tube part of an iron-constantan thermocouple. The constan-

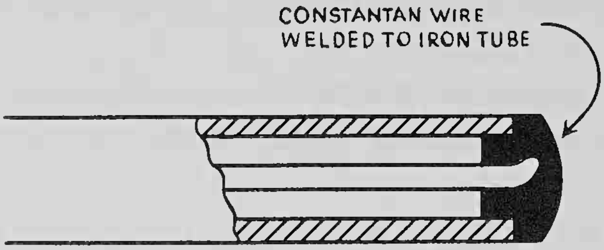


Figure 9–14. Pencil-type thermocouple.

tan element, which mechanically is considerably weaker than iron, is fully enclosed by the iron. This gives minimum mass with relatively good mechanical protection without well. However, easy exchangeability is again sacrificed. Response speed, however, is considerably increased.

A high-speed thermocouple developed by Leeds & Northrup for steam temperature control is illustrated in Figure 9–15. To improve thermal contact between thermocouple and well, spring pressure is applied on the sleeve which forces the silver plug—containing the measuring junction—against the base of the well. The purpose of the silver plug is to assure a large area of contact by a material of low thermal resistance.

A similar approach has been used in Foxboro's "Dynatherm" for a resistance thermometer, as illustrated in Figure 9–16. The temperature -sensitive resistance element is wound on a silver core. Metal foil in the base of the well assists the heat transfer to the solid silver tip and core. Thus the effect of the air space between resistance temperature detector and inside wall of the well is minimized and heat flow is rapidly and evenly distributed through the aluminum foil and the silver parts.

Nothing has been said so far about radiation-type primary elements. Since they do not depend on convection and conduction, the corresponding time-delaying factors do not apply. Only the mass of the radiation-sensitive element (usually a thermocouple or thermopile) must be considered. Time constants are 0.2 to 0.5 sec.

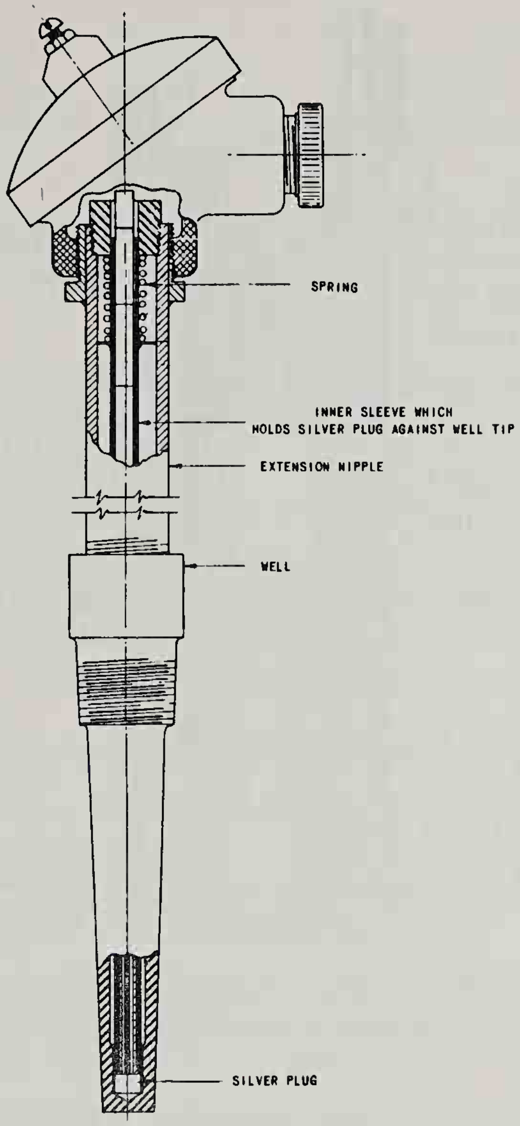


Figure 9–15. High-speed thermocouple. (*Courtesy of Leeds & Northrup*)

Where time constants are relatively long, as in the majority of temperature measurements, the speed of the controller becomes a negligible factor in the control system, and any improvement there is without influence on the whole. In choosing a controller for a temperature-controlled process, it is advantageous to know: (1) the process gain in units change of controlled variable per inch displacement of final control element; (2) the dead time; (3) the load changes that can be expected, in terms of temperature changes that would take place if no control were to be applied; (4) the speed with which these load changes occur; (5) the allowable maximum temperature deviation.

Dead times of various components in the control loop are added.

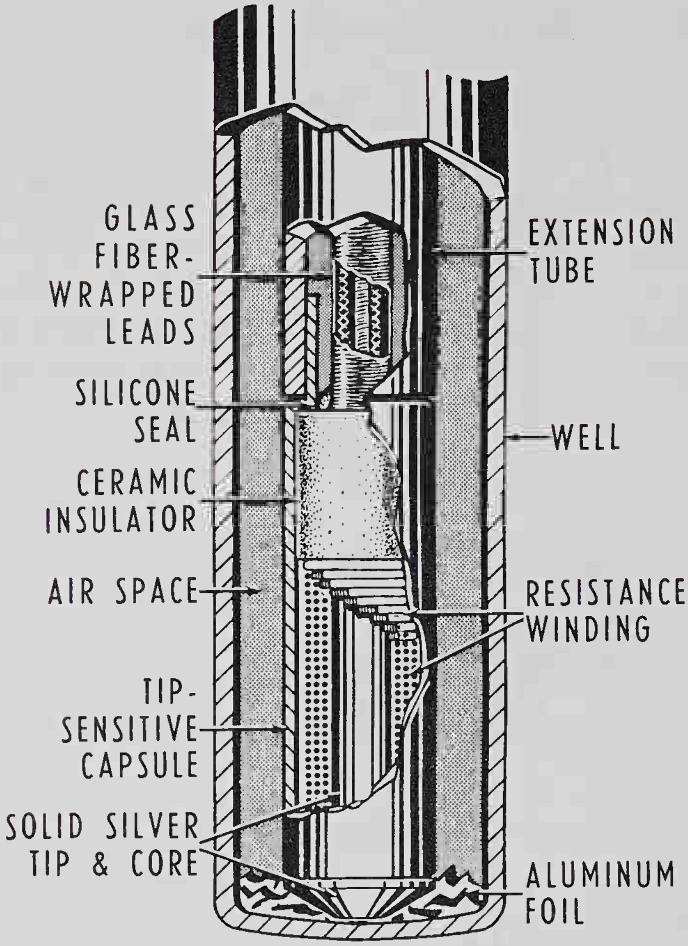


Figure 9–16. "Dynatherm" resistance temperature detector. (Courtesy of Foxboro Co.)

Time constants can be neglected except one, if this one is more than twice the others. It is then possible to make the following decisions:

1. If the load changes are slow, no rate action is required;
2. If the floating rate to be calculated according to page 42 results in a satisfactory answer, this action should be chosen.
3. If proportional-position action as calculated in accordance with the method on page 47 gives a satisfactory answer, this should be chosen; otherwise proportional plus reset plus rate action should be chosen.

This shows that the immediate recourse, as it is frequently taken, to proportional plus reset plus rate action for temperature control, is workable but is certainly not the most economical.

10. CONTROLLERS

The controller links the measuring means and the final control element. The purpose of the controller is to regulate the control point at which the controlled variable is to be maintained, to adjust the proportional speed or band of the signal to the final control element, and to provide responses which may depend not only on the magnitude of a deviation from the controlled variable, but also on the rate of change and the duration of this deviation. How these functions are achieved in different types of controllers is described in the following pages.

Pneumatic Controllers with Position Balance

Pneumatic controllers in which the input is a position differ basically from those where the input is force. As previously pointed out, the position system is primarily used with recorder controllers or indicator controllers.

Set Point Mechanisms. The first function of any controller is to compare the primary feedback from the measuring element with the reference input from the set point mechanism. This is essentially a mathematical operation:

$$\text{(Reference input)} - \text{(Primary feedback)} = \text{Actuating signal}$$

For the purpose of this mathematical operation, mechanical linkages are used like those shown in Figure 10–1. Diagram (a) is self-explanatory, the flapper position is the result of the set point adjustment as well as of the input. In diagram (b) an eccentric disc is mounted on a shaft. The input signal is rotation of the shaft. To change the set point, the relative position of shaft and eccentric disc is adjusted manually by rotating the disc on the shaft. In diagram (c) flapper and nozzle are independently positioned, one by the input, the other by the set point adjustment.

Figure 10–2 illustrates an actual summarizing (or subtracting) linkage as used by the Mason-Neilan Regulator Co. This linkage is carried by the pen mechanism. The input signal from the measuring element is transmitted by the element link to the pen arm. The control pin on the pen connector engages a slot in the control lever and transmits this

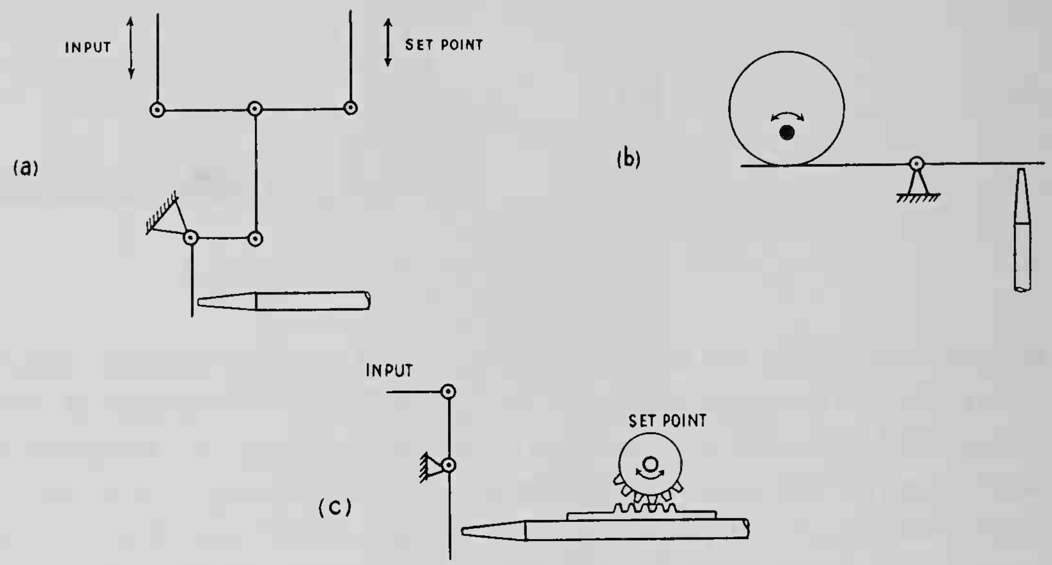


Figure 10–1. Set point mechanisms.

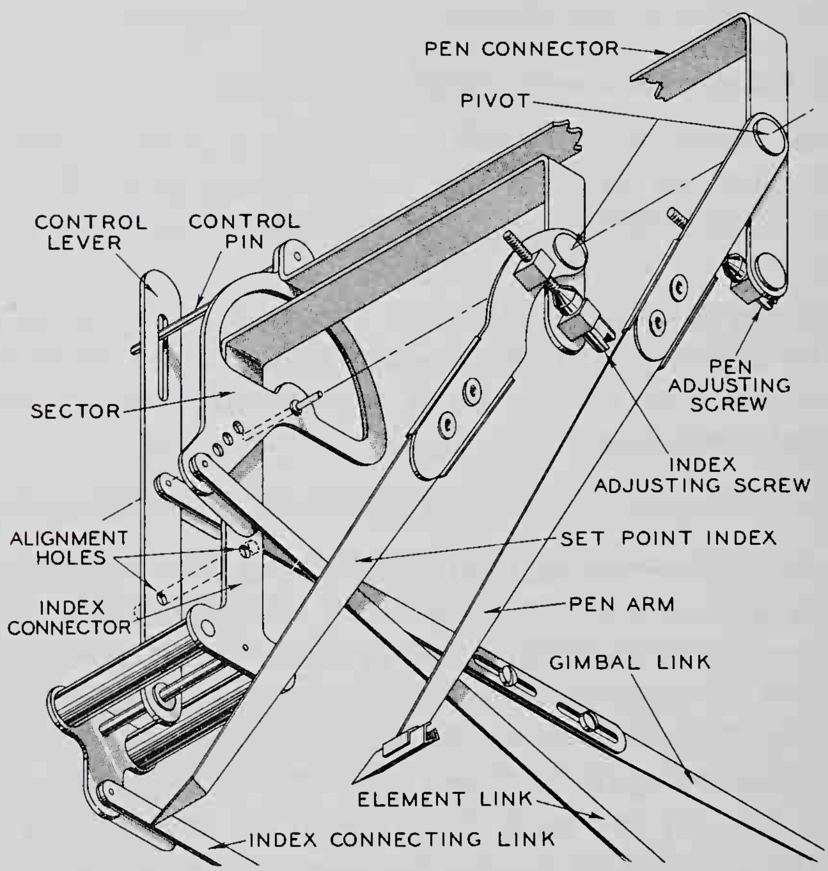


Figure 10–2. Mason-Neilan set point mechanism.

motion to the so-called gimbal link which provides the actuating signal for the flapper.

The index-connecting link transmits the position of the set point adjustment knob to the summarizing linkage. Turning the knob shifts

the position of the gimbal link with respect to the pen. The arrangement is basically that of Figure 10–1a.

Controller Mechanisms. From the set point mechanism the actuating signal is applied to the controller mechanism, where different controller actions are produced. Two-position action is not possible with pneumatic control, unless special relay devices are used. This is because the flapper motion can only produce a gradual change of the nozzle back pressure and not an on-off action. However, since a flapper motion in the order of 0.001 in. usually suffices to change the nozzle back pressure through its full range, arrangements without feedback such as shown in Figure 10–1 can be adjusted to provide proportional bands of one per cent or smaller and may then be considered two-position controllers for most practical purposes.

Flapper-nozzles without feedback are available for proportional bands up to approximately 10 per cent. Beyond this, feedback becomes a requirement. It must be realized that with a total motion in the order of 0.001 in., the flapper is accurately positioned within 0.00001 in. when a change of one per cent in the controller output signal is desired. As a matter of fact, a high degree of accuracy for considerably less motion is actually obtained.

Feedback mechanisms may be divided in those repositioning the flapper and in those repositioning the nozzle. An example of each is described here.

Figure 10–3 illustrates the operation of a proportional-position controller as developed by the Minneapolis-Honeywell Regulator Co. The set point mechanism follows the same principle as the one previously described. The flapper position is determined by the displacement balance between actuating signal and feedback signal. The latter is obtained from the controller output signal through the bellows. The function of this mechanism is similar to the set point mechanism, i.e., a subtraction of two signals. In this case, the equation reads

$$(\text{Actuating signal}) - (\text{Controller output signal}) = \text{Flapper position}$$

The difference with the set point mechanism, however, is that in the equation of the proportional-position mechanism, a closed loop relation exists: the flapper position determines the nozzle back pressure which determines the controller output signal which in turn determines the flapper position. The relationship can be expressed in form of a diagram, as in Figure 10–4. The feedback mechanism is expressed by the summing point in which the controller output signal is subtracted from the actuating signal. The pilot valve amplifies the result and the symbol used is that of an amplifier.

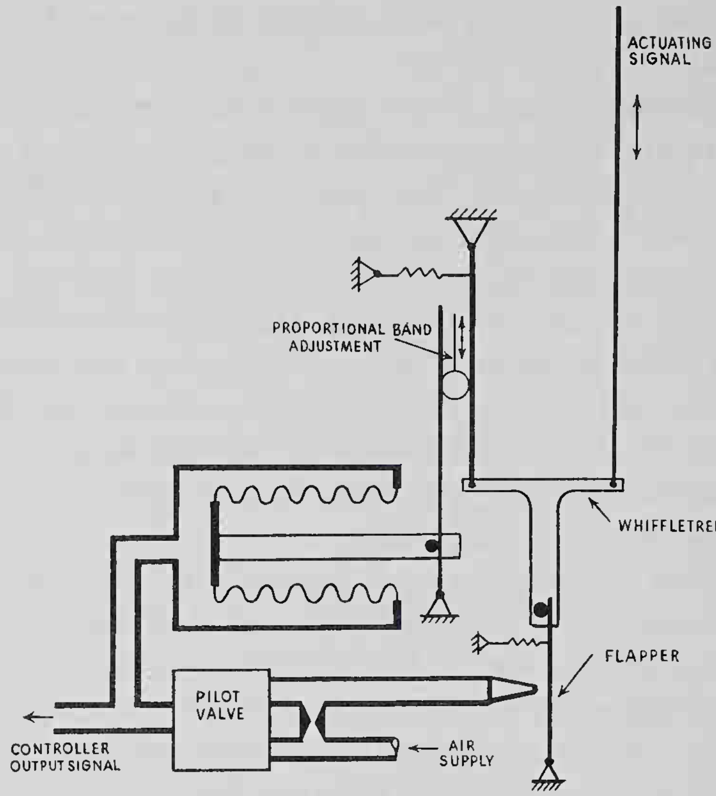


Figure 10–3. Proportional-position controller with flapper feedback.

Going back to Figure 10–3, the proportional band adjustment is indicated by an adjustable fulcrum between two links. This amplifies to a certain extent the feedback motion of the bellows. It is shown as an amplifier in the feedback loop of Figure 10–4; an arrow indicates that the gain of this amplifier, i.e., the proportional band, is adjustable.

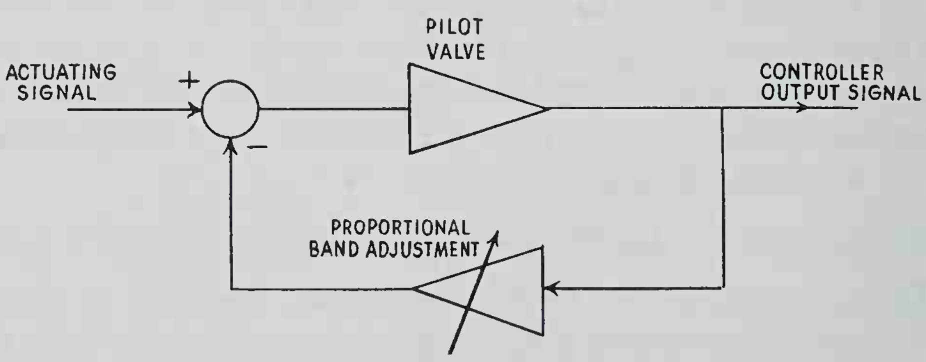


Figure 10–4. Diagram of proportional-position controller.

One other feature is essential in this construction. As seen in Figure 10–3, the flapper is kept in contact with a pin on the whiffletree by means of the spring. This prevents the flapper from being forced against the nozzle with subsequent mechanical damage. When the whiffletree swings out beyond the point at which the flapper contacts the nozzle,

contact between whiffletree and flapper is simply lost and no damage is done.

Figure 10–5 illustrates a somewhat different method of feedback. It follows a principle used by the Taylor Instrument Cos. The feedback

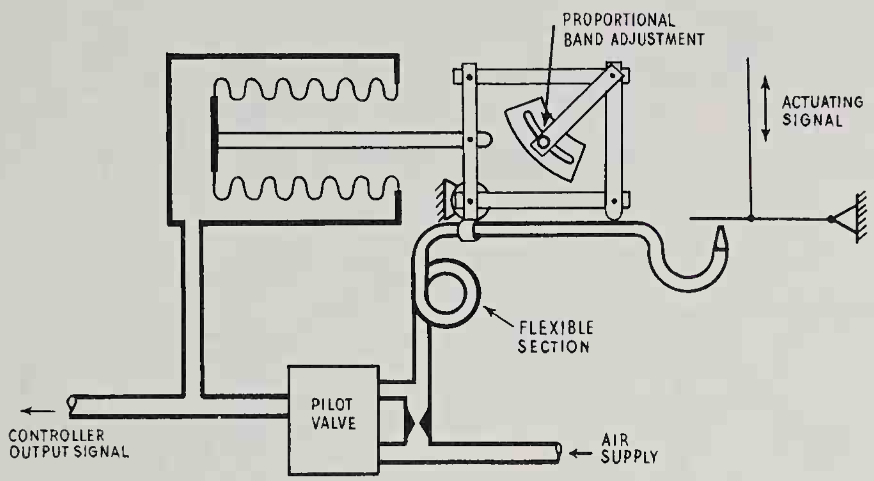


Figure 10–5. Proportional-position controller with nozzle feedback.

operates on the nozzle instead of the baffle as in the previous arrangement. The proportional band is adjusted by means of a parallelogram system of levers which converts the horizontal movement of the bellows into a smaller or larger vertical movement of the nozzle.

The action of the parallelogram is further illustrated in Figure 10–6. The arrangement has two fixed pivots, A and B. Of these pivots, B can be manually shifted and constitutes the proportional band adjustment which, as mentioned before, is equivalent to changing the gain in the feedback. Figure 10–6 illustrates two adjustment positions. The two upper diagrams apply for one, the two lower for the other. By displacing the push rod through a horizontal distance s, the nozzle tip moves vertically, increasing distance d_1 to d_2. In changing the pivot B from B_1 to B_2, and displacing the push rod again through the same distance s, the resulting increase of distance d_1 to d_3 is considerably more than d_1 to d_2. A diagram for this arrangement would be the same as that in Figure 10–4 which can be applied for either the flapper or the nozzle feedback.

Mechanisms for reset and rate action differ from those for proportional-position action because their effect includes a time function. The standard method for obtaining time functions in pneumatic elements consists in combining a resistance and a capacitance, as shown in Figure 10–7. A change of air signal will displace the bellows. However, the response of the bellows is delayed because of the resistance and the capacitance in the

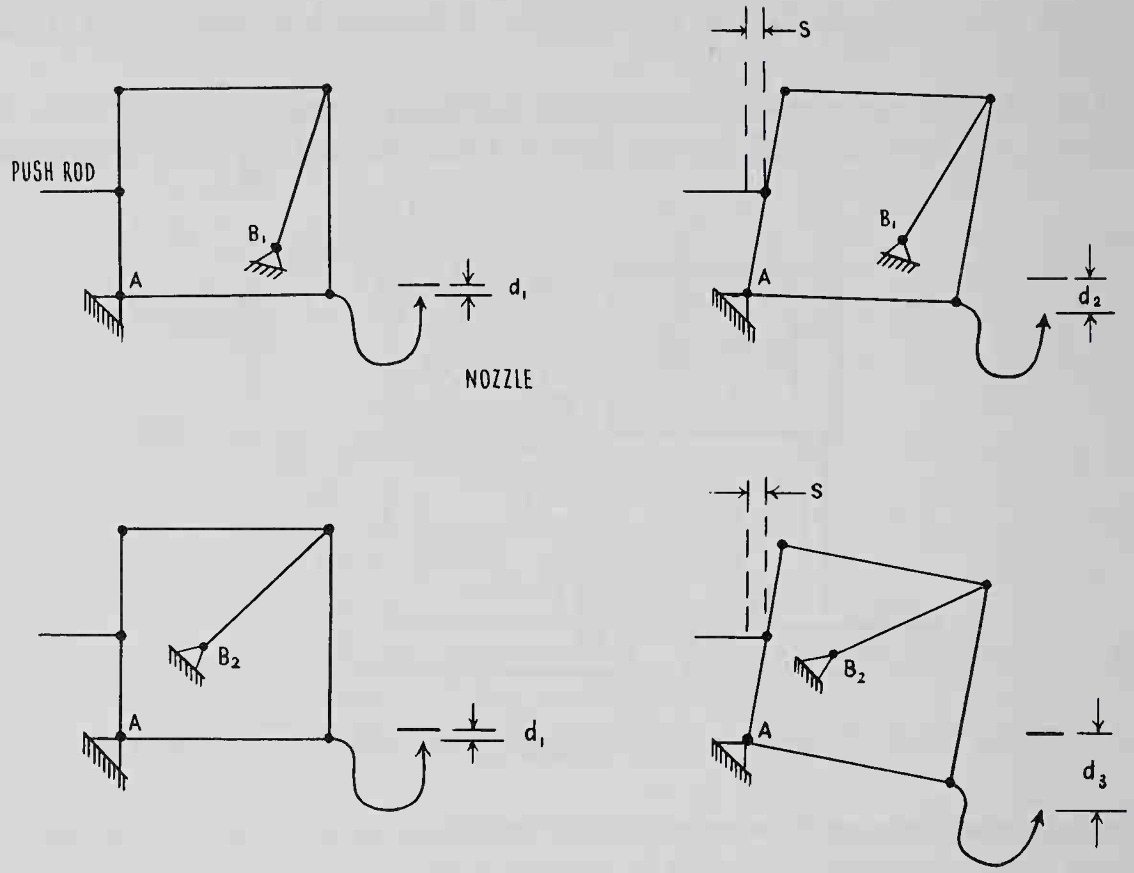


Figure 10–6. Adjustment of proportional band by parallelogram.

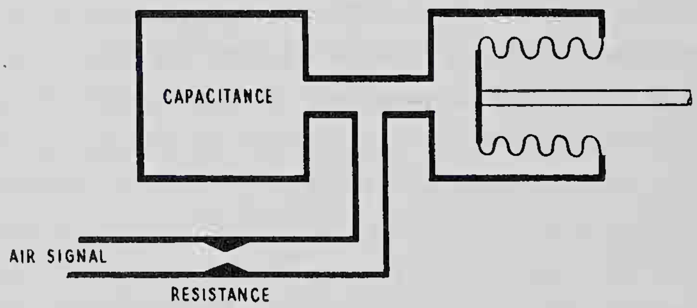


Figure 10–7. Pneumatic *RC* network.

circuit. For example, an increase in pressure of the air signal results in flow across the orifice resistance. This flow continues until the pressures on both sides of the resistance are equalized. The air volume required to equalize the pressures depends on the magnitude of the capacitance. Hence it follows that the time required to transmit the change in air pressure, i.e., the air signal to the bellows, depends on the magnitudes of resistance and capacitance.

The combination of a resistance *R* and a capacitance *C* to produce the described action is called an *RC* circuit. Frequently, the capacitance

in the bellows and connecting lines in combination with the orifice is large enough to provide the effects of an *RC* circuit. In this case the additional capacitance of the volume chamber in Figure 10–7 is not needed.

Air flow through the orifice is roughly determined by the flow equation $Q = kA\sqrt{(p_2 - p_1)p_1}$. This means that the flow rate diminishes as the pressure difference decreases. The result is that the increase of the bellows pressure and hence its movement follows a curve like the one in Figure 10–8. Since the pressure differential becomes increasingly

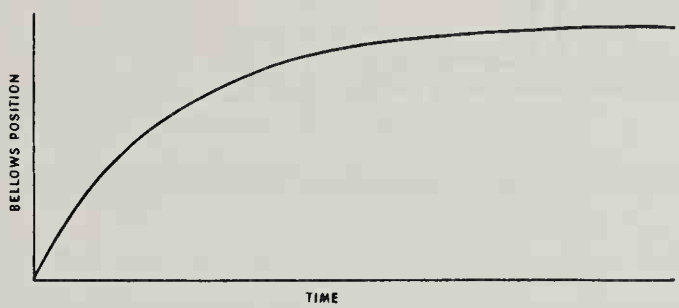


Figure 10–8. Bellows motion in *RC* circuit.

smaller, the flow rate and with it the bellows motion becomes slower. Mathematically, a finite position will never be reached and the curve approaches some finite value continuously without ever reaching it. Under practical conditions, a finite position is obtained. Nevertheless, it would be difficult to give an exact figure for the time it would take the bellows to reach a new position under certain defined conditions. This limitation has led to the usage of the time constant, which has been discussed before. The time constant is the time the bellows needs to change through 63.2 per cent of the maximum motion that results from a signal change.

The method by which proportional-position and reset actions are combined is illustrated in Figure 10–9. Upon a change, e.g., an increase of controller output signal, the proportional position bellows will act as feedback without delay. This will minimize the flapper motion. The effect of the reset bellows will be gradual because of the *RC* network. Its action opposes that of the proportional position bellows, moving the flapper away from the nozzle, thus increasing the controller output signal beyond the first impulse which was controlled by the proportional-position feedback. This means that positive feedback is added to negative feedback. Figure 10–10 is the resulting diagram. The *RC* network is indicated by a separate box. Since the reset feedback that passes through it is subtracted in the measuring point, and the resulting signal

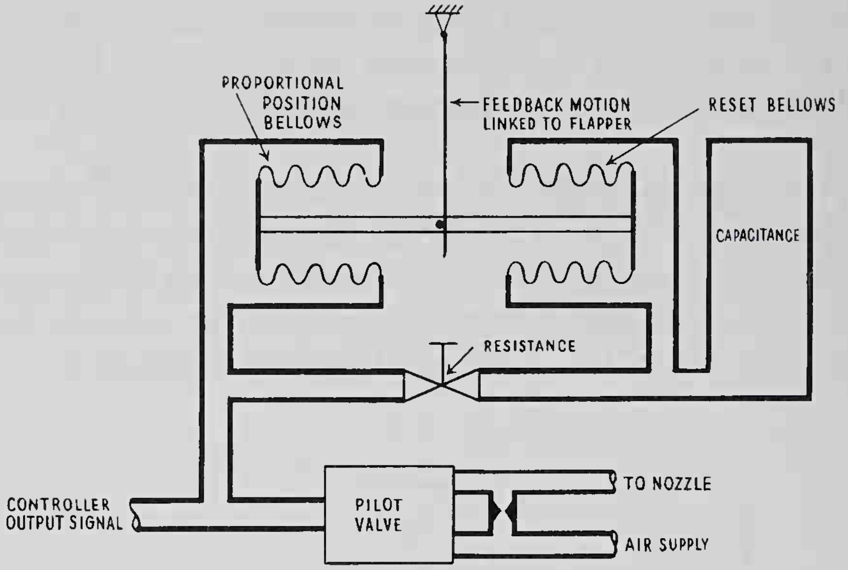


Figure 10–9. Proportional plus reset feedback.

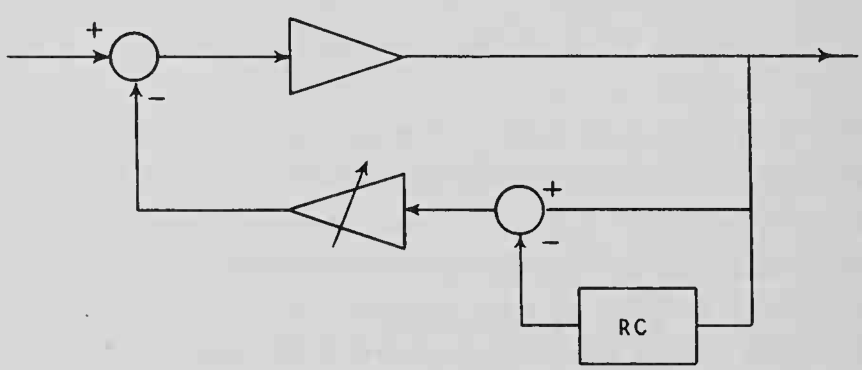


Figure 10–10. Schematic drawing of proportional plus reset controller.

becomes subtracted a second time at the summing point where it is added to the actuating signal, the net effect is that the reset feedback is positive.

The proportional band adjustment affects both feedbacks, proportional and reset. This is necessary because reset action is to repeat the effect of the proportional-position action by a certain number of times per minute, which is expressed in the reset rate. Hence, the effect of the reset action must be proportional to that of proportional action which is assured by a common gain adjustment.

The reset rate as such can best be adjusted by making the resistance variable. This is accomplished either through a needle valve which provides the necessary adjustable restriction or by a capillary, the length of which can be changed.

Rate action implies that the controller output signal be changed with the rate of change of the actuating signal. Rate action is provided by adding another RC arrangement as illustrated in Figure 10–11 (cf. Figure 10–10). The dashed line indicates an alternate connection of essentially

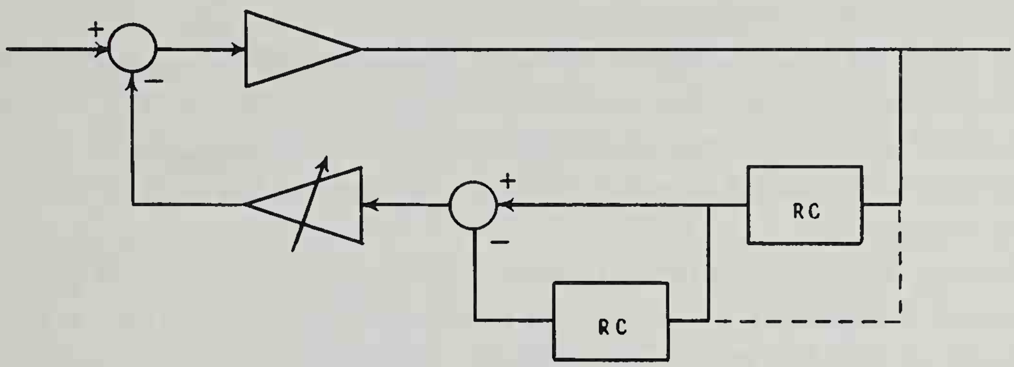


Figure 10–11. Schematic drawing of proportional plus reset plus rate controller.

the same effect. This second RC network delays the feedback action, thereby allowing the actuating signal to pass undiminished through the first summing point. This is further illustrated in Figure 10–12. The bellows opposing the reset bellows is delayed in its action by an addi-

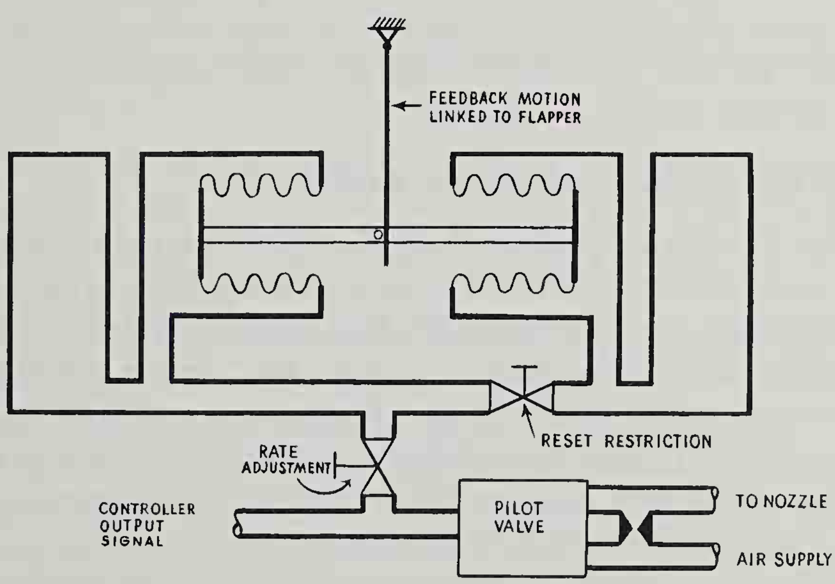


Figure 10–12. Proportional plus reset plus rate feedback.

tional restriction and capacity tank. The consequence is that the proportional-position feedback lags as desired. The magnitude of the lag depends on the magnitude of the controller output signal as well as on the speed with which this change occurs. This results in a larger controller output signal than would be obtained with undelayed propor-

tional-position signal; and this delay will be proportional to the speed with which the actuating signal changes, which is the purpose of rate action.

From the foregoing, it can be concluded that the lag effect of a pneumatic *RC* network produces reset action when used in positive feedback, and rate action in negative feedback.

The effect of reset action is actually the same as if the set point were to be changed. When the controlled variable is below the *actual* set point, the reset action pushes the *effective* set point and with it the *effective* proportional band upward, thereby increasing the corrective action of the controller.

This may result in a difficulty with three-term (proportional plus reset plus rate) controllers when the set point is changed or when a process is started up and the controlled variable is outside the proportional band but rapidly approaching it. As the controlled variable enters the proportional band, the rate action results on a controller output signal which is proportional to the rate of change of the controlled variable. This decelerates the controlled variable and reduces overshooting. Since, however, the *effective* proportional band has changed because of reset action, this effect takes place only when the controlled variable is far within the *actual* proportional band. The result is considerably more overshoot than would be obtained if the rate action could be made independent of the reset action as in other pneumatic controllers described further below.

Pneumatic Controllers with Force Balance

The force balance controller is used where set point and feedback signal are available as calibrated pressure rather than linked with indicator or recording mechanisms. This makes them essentially blind controllers which may be used with or without recorder or indicator. Figure 10–13 shows the essential parts of a Foxboro Consotrol proportional plus reset controller. The outstanding functional part of the unit is the floating disc which acts both as the force-balance detector and as the flapper of a conventional flapper-nozzle system.

Each of the four bellows exerts an upward force on the disc. The net effect of all forces acting simultaneously is the establishment of the horizontal position of the disc, its nearness to the nozzle, and hence the controller output signal.

In operation, any change in pressure in either the primary feedback or set point bellows slightly raises or lowers the corresponding side of the floating disc and causes a change in nozzle back pressure which operates the pilot valve to increase or decrease the controller output sig-

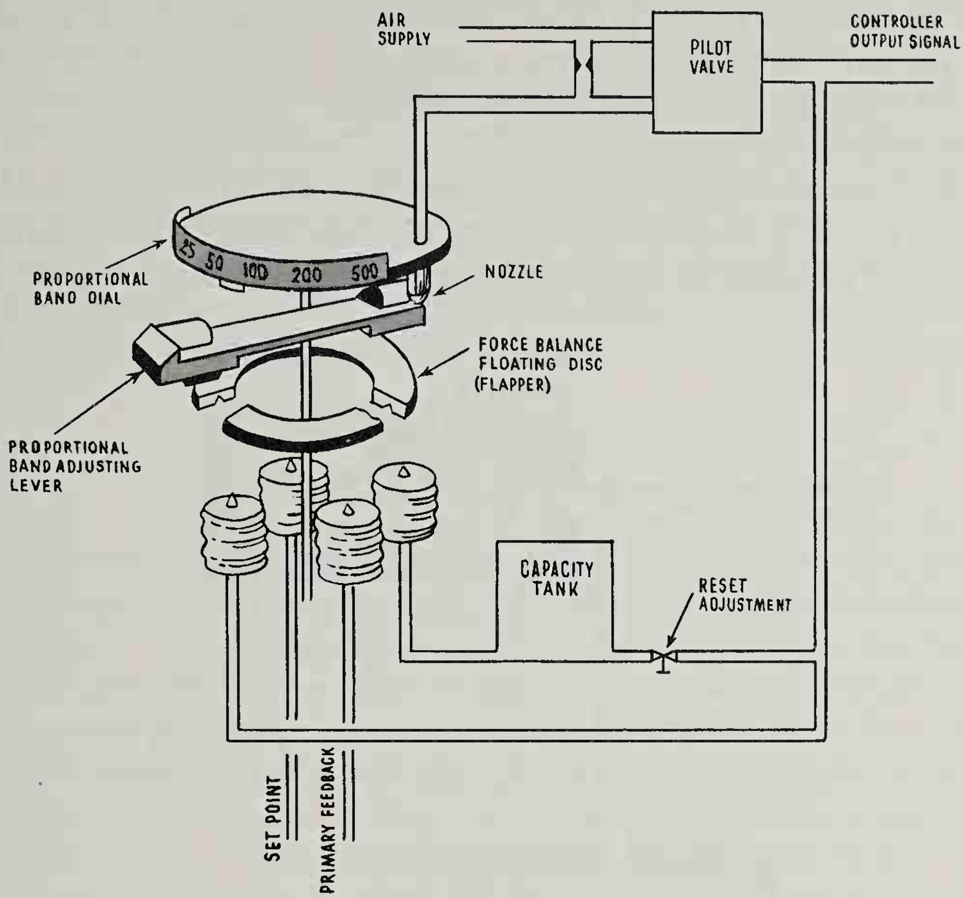


Figure 10–13. Principle of Foxboro Consotrol proportional plus reset controller.

nal. The feedback of this signal acting on the proportional bellows re-establishes a balance of the forces on the disc. The change of controller output signal is thus proportional to primary feedback or set point pressure change.

If there is a sustained differential between set point and primary feedback, a sustained difference between proportional and reset bellows pressures will result. Air will flow from one to the other, causing the disc to alter the nozzle opening and to continuously change the controller output signal to maintain this difference. This is the reset action.

The force balance floating disc rests against the proportional band adjusting lever which provides an adjustable fulcrum about which the floating disc moves. If the position of the proportional band adjusting lever-fulcrum is changed, the moment arm of each bellows about the fulcrum becomes also changed and a different set of forces is required to balance the disc. For example, when the lever is set so that the distance of the proportional bellows from the fulcrum axis is twice that of the primary feedback bellows, a two-psi primary feedback change is balanced by a one-psi output change. A proportional band of 200 per cent is the result.

When the distance of the proportional bellows from the axis is half that of the primary feedback bellows, a one-psi primary feedback change is balanced by a two-psi output change, which is equivalent to a 50 per cent proportional band.

In 1947 the first stack-type controller was introduced by the Moore Products Co. It was a radical departure from existing pneumatic controllers and resulted in an entirely new design concept. Figure 10–14

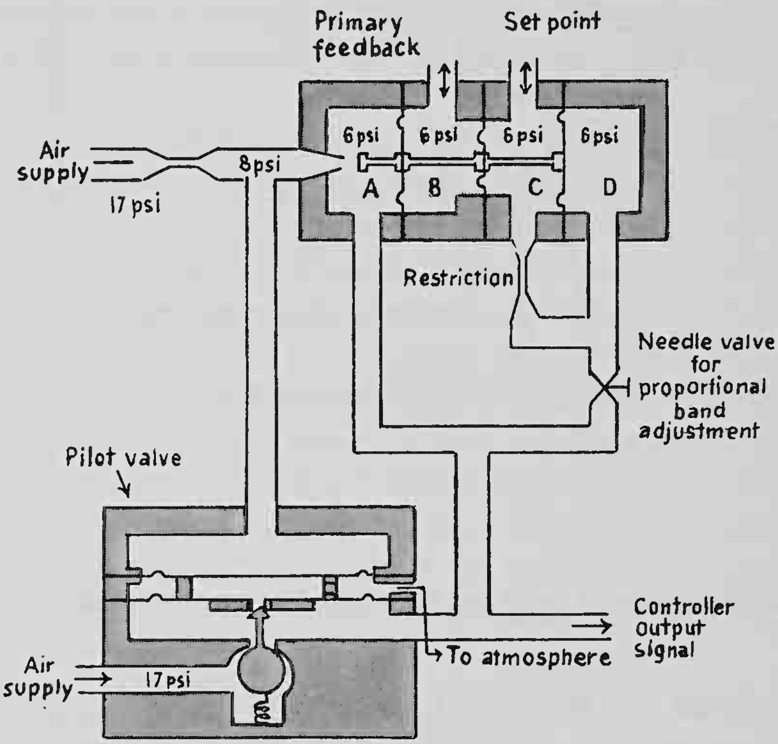


Figure 10–14. Principle of Moore proportional-position controller.

shows the principle of the Moore proportional controller together with the pilot valve which becomes an inherent part of the control circuit. The upper diaphragm of the pilot valve is of smaller area than the lower. Therefore, in order to balance the forces across the two-diaphragm assembly, the pressure on the upper diaphragm must exceed that under the lower diaphragm by a certain amount. These pressures correspond to nozzle back pressure and controller output signal, respectively. Suppose an excess of 2 psi nozzle back pressure suffices to balance the controller output signal. This means that the balanced pilot valve maintains a 2 psi pressure difference between controller output signal and nozzle back pressure. The nozzle bleeds directly into the controller output. The purpose of this arrangement—to approach null-balance conditions—has been described in a previous chapter.

The flapper is part of a rigid stem connected to three diaphragms.

When any of these diaphragms deflect, the whole stack including the flapper is bound to move.

The diaphragm between chambers B and C has half the area of either of the other two diaphragms in the stack. The pressure in chamber B, for example, exerts force in two directions: one through the left diaphragm and one through the right. Since the area is twice as much on its left, the force in this direction is also twice that of the force to the right. By considering forces to the right positive and those to the left negative, the following equation for balanced conditions can be applied:

$$A - B + 0.5B - 0.5C + C - D = 0$$

or simply

$$A - 0.5B + 0.5C - D = 0$$

Since the areas of A, B, C, and D are equal, they cancel out, and it is only necessary to insert the pressures that exist in the respective chambers.

Figure 10–14 shows certain air pressure values which are assumed to represent initial conditions. Suppose the primary feedback signal changes and increases the pressure in chamber B to 10 psi. The increased air pressure exerts a force toward the left and the flapper moves toward the nozzle. This raises the nozzle back pressure and as a consequence the pilot valve acts to increase the pressure of the controller output signal.

The increased output pressure is fed back into chamber A. If the needle valve of the proportional band adjustment is closed, the pressure in D will remain constant.

In order to balance for the new condition it is necessary that

$$A = 0.5B - 0.5C + D$$

and inserting values,

$$A = 5 - 3 + 6 = 8$$

This means that the controller output signal increases by two psi for four psi change in primary feedback signal, which corresponds with a proportional band of 200 per cent.

To increase the gain of the controller and obtain, e.g., twice the response in the controller output signal for a given change in primary feedback, i.e., a proportional band of 50 per cent, the needle valve of the proportional band adjustment is slightly opened.

This changes the behavior of the controller because the feedback, which is directly applied to the A chamber, is now also active in the D chamber, although less because the needle valve reduces the air pressure. As air flows through the needle valve, increasing the pressure in the D chamber, air also flows from D to C through the restriction shown. This will not increase the pressure in C since the set point pressure is con-

trolled by a regulator which, once it is set, will keep the pressure constant. The result is that the pressure in D is maintained between the controller output signal and the set point pressure. Once these two pressures are given, pressure in D depends on the amount of opening of the needle valve.

If the needle valve is positioned so that the increase from 6 to 10 psi in controller output signal raises the pressure in chamber D to 9 psi, then a 4:3 relation exists between the two pressure increments. For a 4 psi change in primary feedback as in the above example, it is then necessary to change the controller output signal to 14 psi in order to obtain balance. This will raise the pressure in chamber D to 12 psi, and inserting the values in the above equation will confirm the balance.

Figure 10–15 adds reset action to the mechanism illustrated in Figure 10–14. The connection between chambers C and D is taken out. To main-

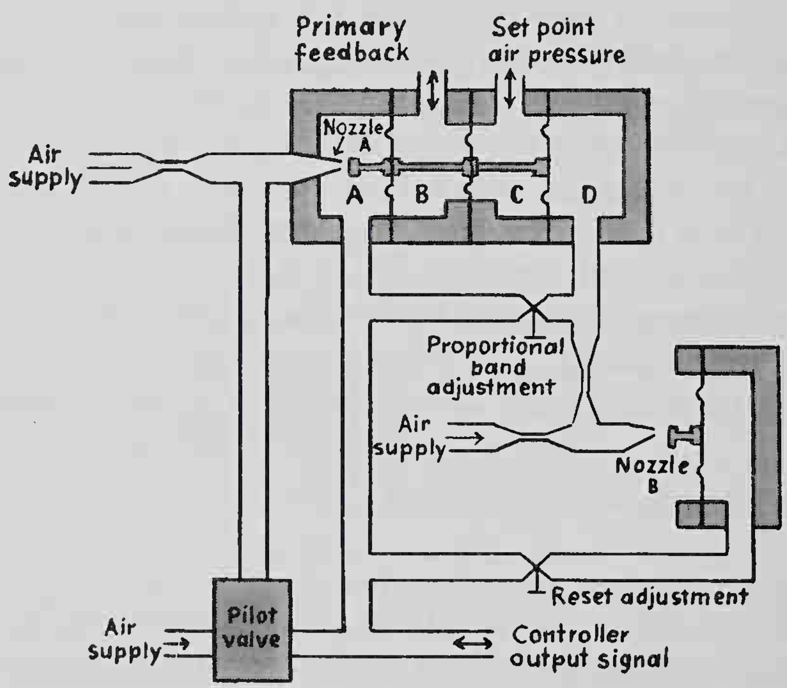


Figure 10–15. Principle of Moore proportional plus reset controller.

tain the feedback pressure in D a second nozzle is added. Chamber D connects through a restriction to nozzle B.

The back pressure in the B nozzle under otherwise static conditions is controlled by a flapper rigidly connected to a diaphragm. The back chamber of the diaphragm constitutes the capacitance of the reset action. Air which flows through the needle valve of the reset adjustment would gradually flex the diaphragm, moving the flapper toward the B nozzle and thereby increasing the back pressure. Closing the reset

adjustment when the nozzle is wide open maintains a minimum back pressure.

The air pressure in the D chamber assumes a value between the controller output signal, i.e., the A chamber pressure, and the B nozzle back pressure. The wider open the proportional band adjustment is, the closer will be this value to the controller output signal and, hence, the narrower will be the proportional band.

At this point, opening the reset adjustment by a certain amount results in a gradual build-up of pressure behind the diaphragm, moving the flapper toward the B nozzle. This increases the pressure also in chamber D which in turn closes nozzle A, and consequently raises the controller output signal. A gradual increase of controller output signal is thus produced as long as an unbalance exists. The rate of this increase is determined by the amount of opening of the reset adjustment.

Figure 10–16 shows how rate action is added to the schema of Figure 10–15. The unit is inserted between the primary feedback and the propor-

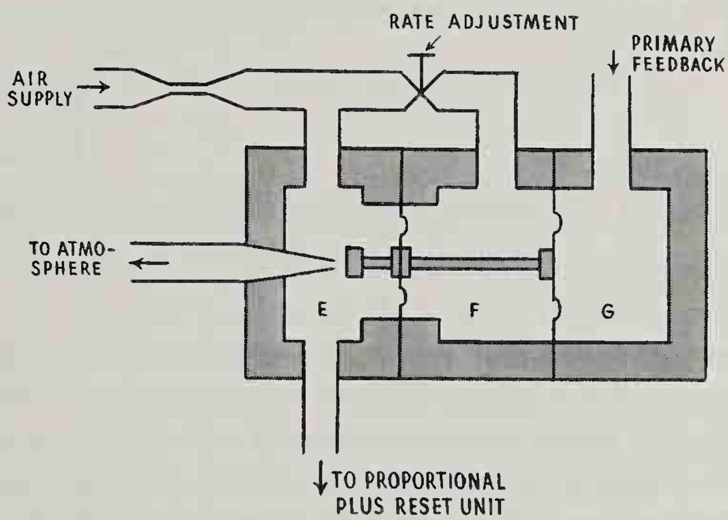


Figure 10–16. Rate action assembly for Figure 10–15.

tional-plus-reset assembly. The pressure of chamber E is applied to chamber B of Figure 10–15. Air is admitted to chamber E through a restriction and flows out to atmosphere through a nozzle. As the outlet area through the nozzle is reduced by the approaching flapper, the pressure in E increases. Assuming that the diaphragm between E and F is about 0.2 times the area of the diaphragm between F and G, then

$$0.2E - 0.2F + F = G$$

or

$$0.2E + 0.8F = G$$

In balanced conditions $E = F$ and therefore $E = F = G$. When the pri-

mary feedback is 6 psi initially and increases to 10 psi, the pressure in E will also rise because the flapper moves toward the nozzle. The pressure rise in F is delayed because of the restriction and capacitance in the chamber. Initially, it is 6 psi. The pressure in E would have to rise to 26 psi in order to effect balance under these conditions, and would therefore amplify the primary feedback signal considerably. Gradually, however, the pressure in F also rises and the pressure in E is reduced accordingly. The amplification in chamber E depends on the rate adjustment and also how fast the primary feedback signal changes.

Figure 10–17 is the diagram of this arrangement. Comparing this with the three-term controller of Figure 10–11 shows a significant difference;

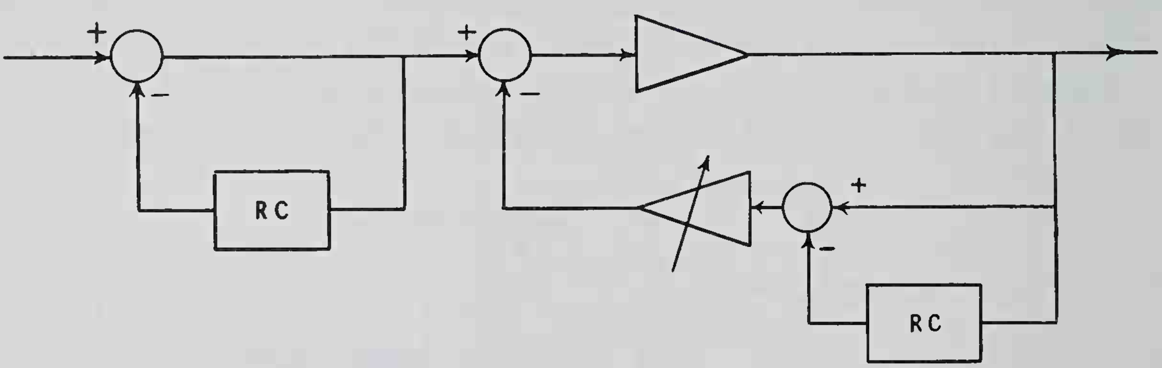


Figure 10–17. Schematic drawing of Moore proportional plus reset plus rate controller.

the rate action is now completely separated from the proportional plus reset mechanism. The result is that the shifting of the effective proportional band due to reset action is without effect on rate action. When the controlled variable is outside the actual proportional band, rate action will dampen the process as soon as the controlled variable enters the actual proportional band of the controller.

The Moore three-term controller is not built in separate units as may be implied from the preceding description. Actually all these units including the pilot valve are combined in one cylindrical assembly measuring 3.375 in. in diameter and about 11 in. in length.

Hydraulic Controllers

Pilot valves and flapper-nozzles as used in pneumatic controls are rarely used in hydraulic controls; four-way valves and jet pipes are generally preferred. Since the output from the jet pipe has considerable signal range both in capacity and pressure it can be frequently used without a four-way valve. Figure 10–18 shows however direct coupling of the two

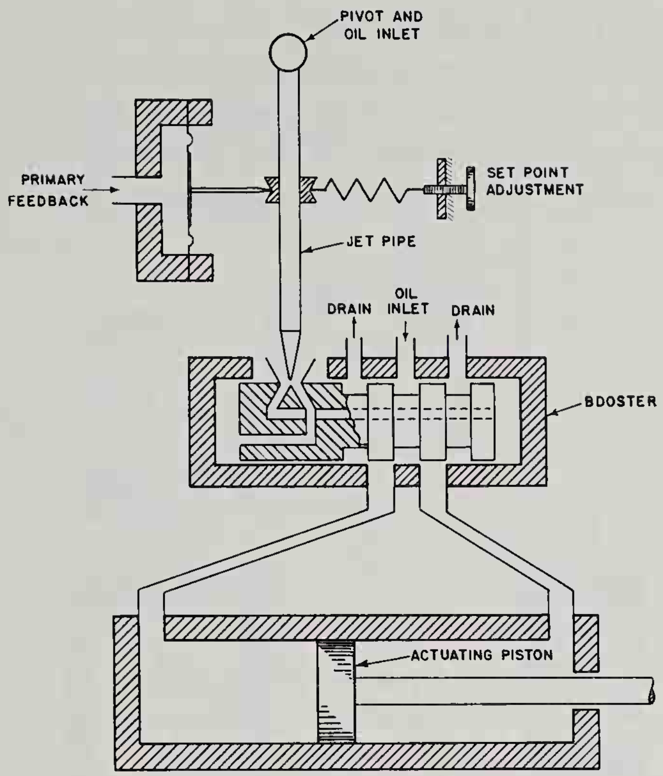


Figure 10–18. Hydraulic floating controller.

components in order to amplify the jet pipe signal. This combination of jet pipe and four-way valve is called a *booster.*

The primary feedback is applied directly to the jet pipe by means of a diaphragm. It is counter-balanced by a compression spring. Thus the actuating signal—the deflection of the jet pipe—is the result of two opposing forces: the primary feedback and the set point spring. The set point adjustment simply changes the compression of the spring.

When the primary feedback increases, the jet pipe swings counter-clockwise. This directs oil to the orifice which opens to the left back chamber of the booster spool. The resulting spool motion, moving toward the right, tends to re-center the two orifices under the jet pipe. When this is achieved, pressure in the left and right back chambers are equal and the spool motion stops. In other words, the spool faithfully follows the motion of the jet pipe. For the rest, the booster action is that of any four-way valve. When the spool is displaced to the right, oil flows into the right chamber of the actuating cylinder. The resulting motion of the actuating piston continues as long as the jet pipe is displaced from its mid-position. The larger the deflection, the wider open are the ports of the four-way valve, and the faster is the actuating piston motion. The action is therefore proportional-speed floating.

Hydraulic controllers frequently use proportional-speed floating

action, but by means of feedback they are readily converted into proportional-position controllers. This is illustrated in Figure 10–19. The primary feedback is applied to a bellows. The bellows deflection is balanced by the spring characteristic of the bellows itself and a tension spring. The set point is adjusted by changing the tension.

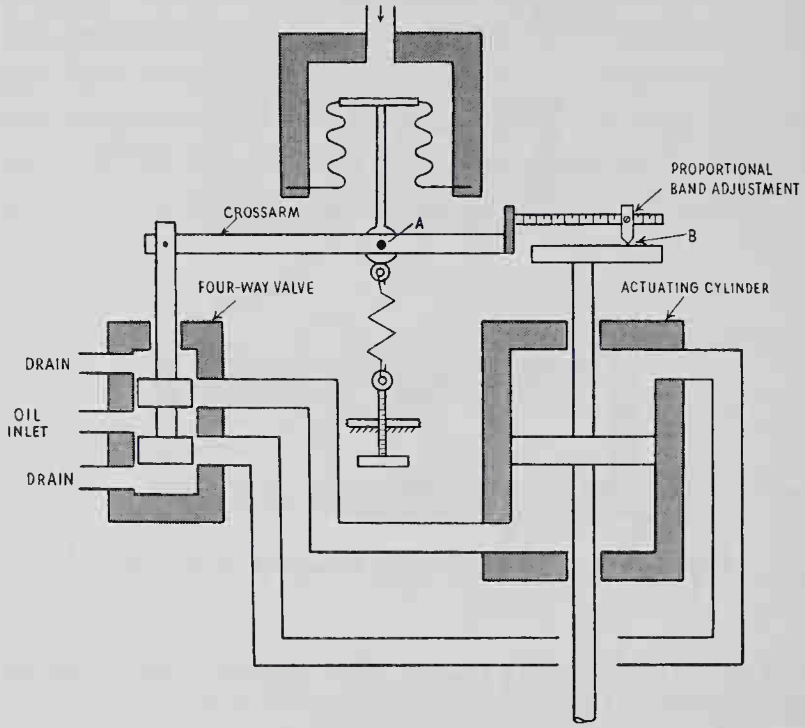


Figure 10–19. Hydraulic proportional-position controller.

The crossarm pivots about *B* when the bellows moves, and about *A* when the actuating piston moves. A change in signal displaces the bellows, rotating the crossarm about *B* and positioning the spool of the four-way valve. This admits oil to one side of the actuating piston. The resulting motion is fed back to the crossarm and, by rotation about *A*, to the spool which is thereby recentered, stopping further motion of the spool. Thus a position is obtained which is proportional to the primary feedback signal. The proportion of position to signal, i.e., the proportional band, is determined by the position of pivot *B*.

Although the hydraulic controller is preferred for proportional-floating action and is readily adaptable for proportional-position action, the combination of the two into proportional plus reset control is somewhat more difficult. The main cause is the incompressibility of the fluid. The resistance-capacitance network of the pneumatic circuit is not readily applicable under these conditions. Pressure change by compression does not apply for liquids. The method used for a hydraulic proportional plus reset controller is shown in Figure 10–20.

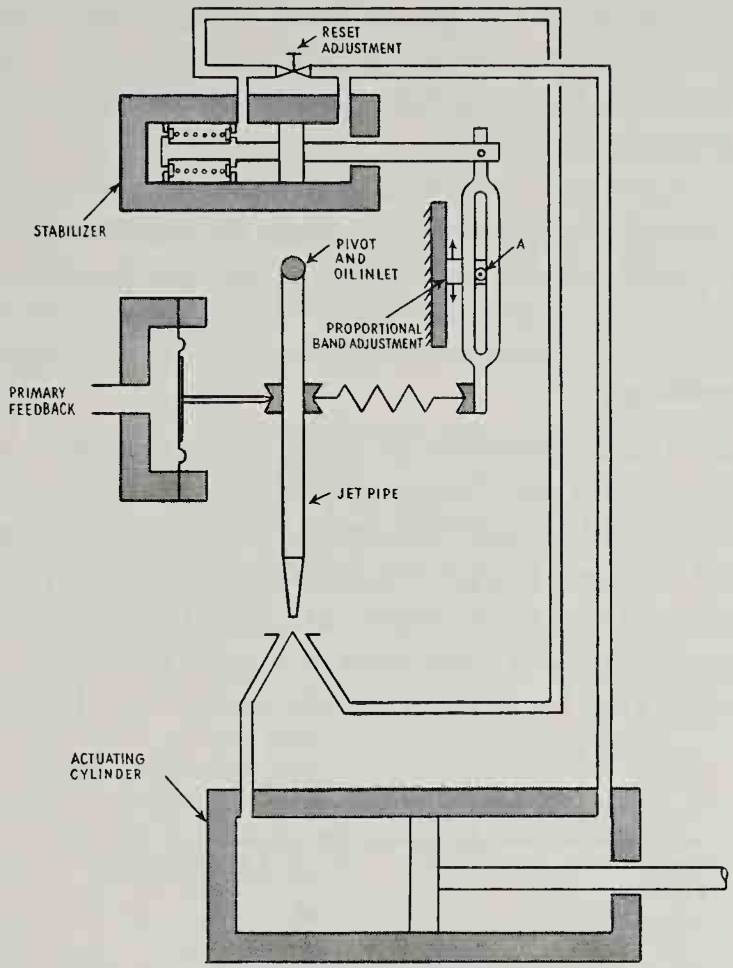


Figure 10–20. Hydraulic proportional plus reset controller.

Control actions are obtained by the hydraulic stabilizer. Suppose the primary feedback signal rises. This moves the jet pipe counterclockwise. The resulting increase of pressure in the corresponding receiving orifice will drive the piston in the stabilizer to the right. The small bypass flow through the reset adjustment may be disregarded here. The oil volume which is displaced from the right side of the stabilizer piston flows into the actuating cylinder. The actuating piston thus assumes a position which is proportional to the stabilizer position.

The movement of the stabilizer piston feeds back to the jet pipe by means of a lever—rotating about pivot A—and a spring, rebalancing the jet pipe. Thus proportional-position action is obtained. The proportional band is adjusted by shifting pivot A. The piston of the actuating cylinder is positioned in proportion to the primary feedback signal.

Displacement of the stabilizer from its middle position compresses the spring in the stabilizer toward the left or right, according to the direction of the displacement. In any case the spring force will tend to return the stabilizer piston to its center position. The volume which is displaced

in this motion must flow through the reset adjustment, which is a needle valve. Hence the amount of opening of the needle valve, i.e., the adjustment of the reset rate, determines the speed with which the stabilizer piston returns to its center position.

The result of the return motion of the stabilizer piston is that, through the feedback linkage, the jet pipe continues by reset action the motion of the actuating piston which was initiated by the proportional-position action.

Electric Controllers

The term electric controllers, as used here, includes electronic controllers. The simpler forms of electric control are most widely used in air conditioning and heating control systems. However, they are not necessarily restricted to this use and their inherent simplicity and proven reliability should entitle them to many additional control tasks.

A two-position control is illustrated in Figure 10–21. The primary contacts may be operated by various means such as the bimetallic element

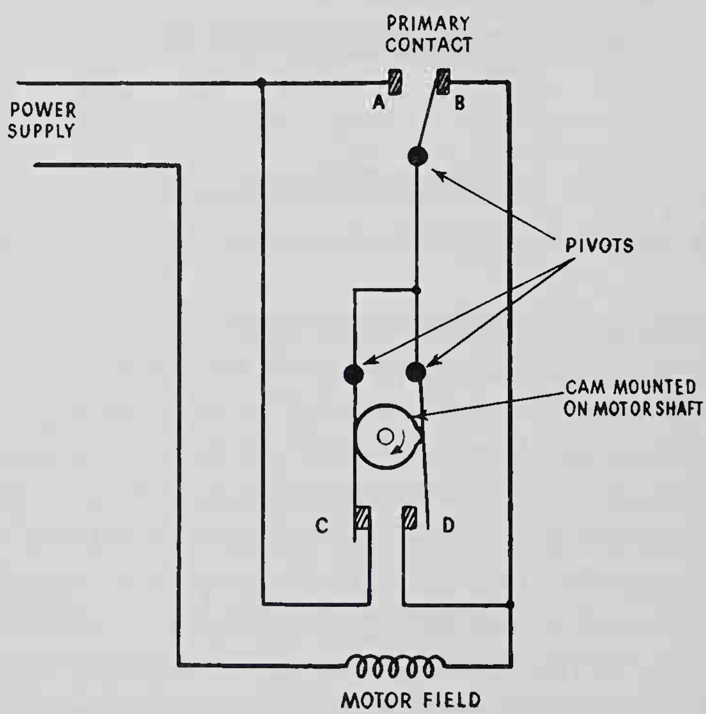


Figure 10–21. Electric two-position controller.

of a thermostat or the electrodes of a level controller, as was shown in Figure 4–3. In the position shown, primary contact *B* just closed and the field winding of the motor is energized through contacts *C* and *B*. As soon as the motor starts rotating, the cam mounted on its shaft will close contact *D*. The field winding remains energized through contacts *C* and *D*

independent of the action of the primary contacts until the motor has rotated through 180 degrees. At this point, the cam opens contact *C* and the motor stops. It now requires closure of the *A* contact to again energize the motor and drive it through another 180 degrees. The motor action is used to position a final control element.

A typical electric single-speed floating controller was described before (Figure 4–8); the same type of shaded-pole reversible single phase motor may be used for proportional control action as illustrated in Figure 10–22. This is the principle of the Honeywell Series 90 circuit. In this

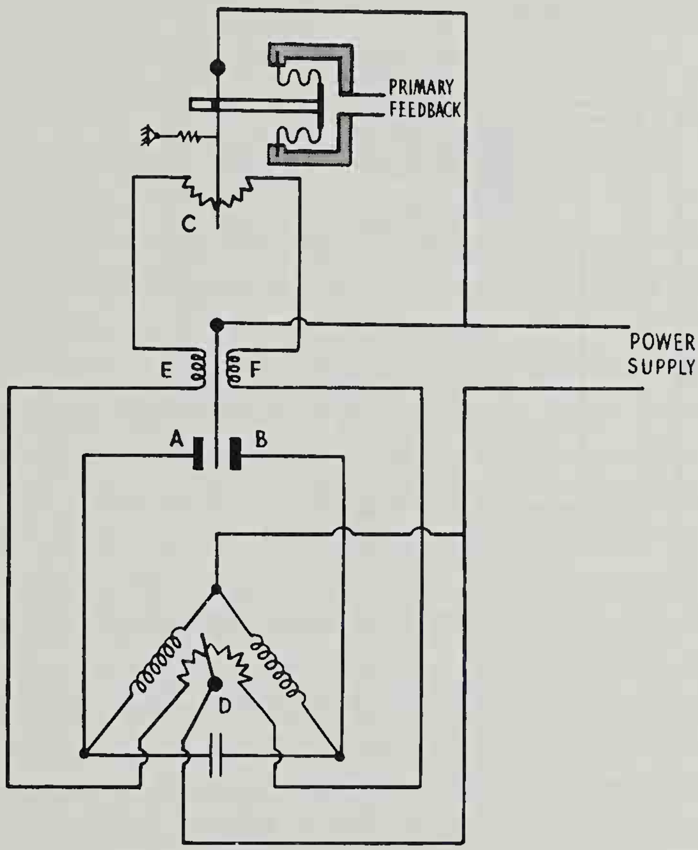


Figure 10–22. Electric proportional controller.

arrangement, the primary feedback is applied to a bellows or equivalent. The bellows positions the sliding contact of slidewire *C*. A relay is built into the motor unit. It consists of relay coils *E* and *F*, and contacts *A* and *B*. If more current flows through relay coil *E* than through *F*, contact *A* will close. Conversely, if more current flows through *F* than through *E*, contact *B* will close. The motor will move in one direction or the other, depending upon which contact is closed. The sliding contact of a second potentiometer, *D*, is mounted on the shaft of the motor. In the condition shown, both potentiometers, *C* and *D*, are in such positions that exactly the same amount of current flows through relay coils

E and *F*. When the primary feedback signal increases the sliding contact of *C* moves clockwise. This increases the current through *E* and decreases the current through *F*. Consequently, contact *A* closes and the motor starts rotating in clockwise direction. This moves the sliding contact of *D* until currents through *E* and *F* are again the same and the relay contacts are broken. Thus the amount of motor movement is proportional to the primary feedback signal.

A somewhat different arrangement, using the same motor, is illustrated in Figure 10–23. This has the advantage over the previous one in providing a null balance at the measuring element. The primary feedback is applied to a diaphragm. The resulting force is balanced by a moving coil. Any unbalance results in a deflection, closing either contacts *A* or *B*. This energizes the positioning motor which in rotating positions the sliding contact mounted on its shaft. The resulting change in voltage applied to the moving coil rebalances the forces and opens the circuit.

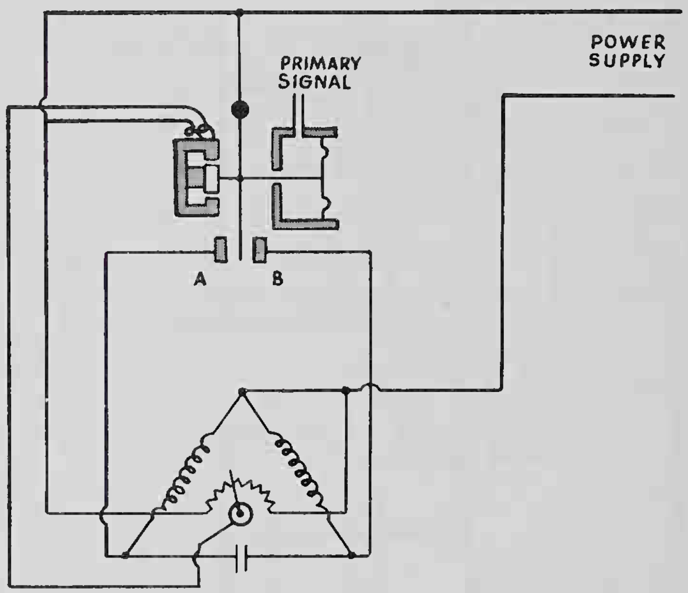


Figure 10–23. Electric proportional controller with null balance.

In order to avoid contacts and slidewires, a number of other arrangements are available. Generally, they belong in a higher price class. Typical for these are synchros (see Figure 7–13) which operate through amplifiers—electronic or magnetic—and servos to obtain proportional-position action.

Where reset and rate actions are required, combinations of capacitances and resistances as illustrated in Figure 10–24 are generally used. The lag network operates by the same characteristics as its pneumatic counterpart (Figure 10–7). Under balanced conditions, the voltage drops across *R* and the load resistance add up to the d.c. supply voltage *E*. Suppose

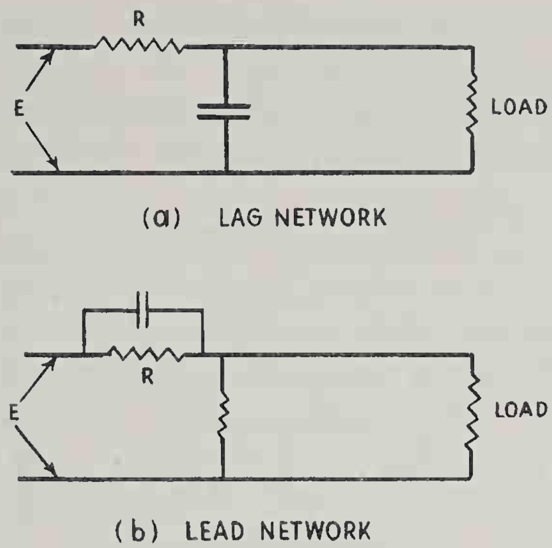


Figure 10–24. Electric *RC* networks.

the supply voltage (Figure 10–24a) decreases. Without the capacitor, the voltages across the resistances would drop correspondingly. With the capacitor, however, an additional discharge current will flow through the load resistance, which diminishes gradually until the capacitor charge corresponds with the new value of *E*. Thus the capacitor has a delaying effect in establishing balanced conditions after a change. It is therefore called a lag network. The situation is the same whether the supply voltage increases or decreases. A curve showing the change of the current through the load after a change in supply voltage would have the same contour as the curve of the bellows motion in a pneumatic *RC* circuit shown in Figure 10–8. The magnitude of the time constant is proportional to the product of capacitance and resistance. A condition of this arrangement is that d.c. voltage is required.

The lead network shown in Figure 10–24b is another method of combining electrical resistance and capacitance. Again under balanced conditions, the circuit behaves the same with or without the capacitor. When the supply voltage *E* changes, however, the capacitor becomes in effect a conductor, shunting out resistor *R*. As the capacitor readjusts its charge, its effect gradually diminishes and finally current will again flow only through *R*. The result is that a change in voltage produces initially a much more pronounced change of voltage across the load resistance, which gradually subsides and finally assumes the proportion that is given by the resistance values. The voltage across the load resistance actually leads to the supply voltage change and is proportional to its rate of change. The lead network provides rate action in the forward path. If used for negative feedback, it provides reset action. This is the reverse of the pneumatic *RC* circuit and the lag network shown in Figure

10–24a, which provide reset action in positive feedback and rate action in negative feedback.

Figure 10–25 illustrates the use of *RC* networks in a proportional plus reset plus rate controller as built by Leeds & Northrup. Two Wheatstone

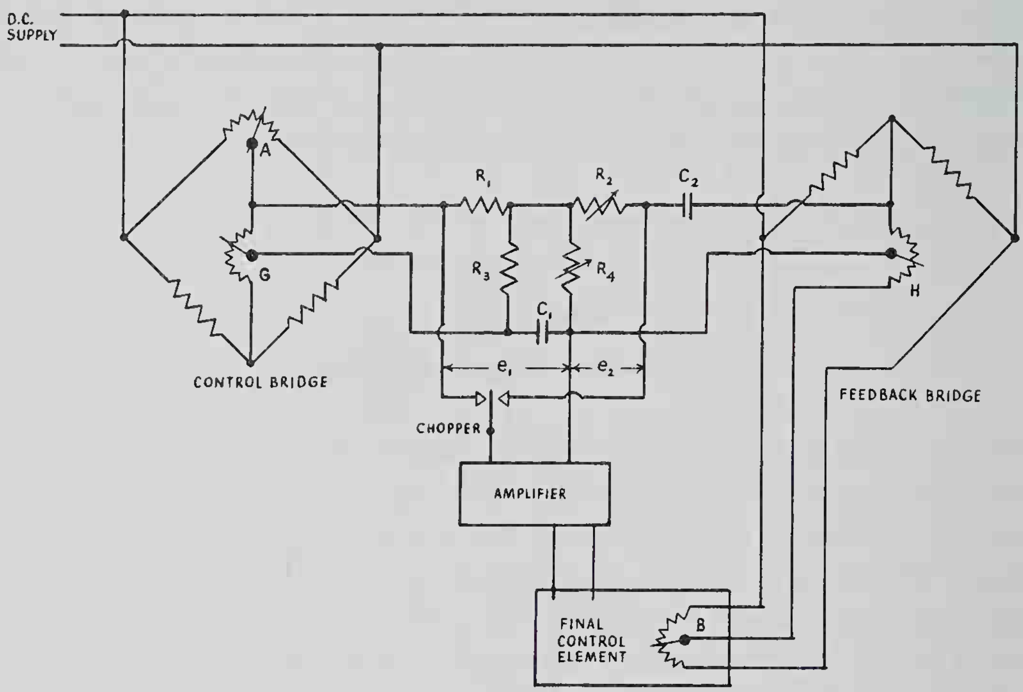


Figure 10–25. Principle of Leeds & Northrup proportional plus reset plus rate controller.

bridges—a control bridge and a feedback bridge—are combined in this arrangement. The sliding contact of slidewire *A* is mechanically linked with the pen arm in the recorder. The set point is adjusted mechanically.

Deviation from the set point moves the sliding contact at *A*, unbalancing the control bridge. This results in a voltage e_1 due to the current flow through the measuring leg of the bridge. Since, however, resistors R_1, R_3, and R_4 and capacitor C_1 constitute a lead network, rate action is provided. Rate time is adjusted by changing the variable resistance R_4.

A chopper, essentially a vibrating reed, rapidly switches from e_1 to e_2, applying to the amplifier either voltage in rapid pulses. As long as a difference between e_1 and e_2 persists, the amplifier puts out a signal to the final control element which is coupled with potentiometer *B*.

Motion of the final control element and hence of the sliding contact at *B* unbalances the feedback bridge. This produces a voltage across capacitor C_2 and resistors R_2 and R_4. As long as the controlled variable does not return to the set point, an e_1 voltage other than zero is maintained. To provide an equal e_2 voltage, a current must flow through re-

sistor R_2, charging capacitor C_2. As C_2 charges, current in the feedback bridge tends to diminish. To keep charging current flowing, the final control element must continue to move, producing a corresponding change in position of the sliding contact at B, until e_1 is again at zero with the controlled variable at the set point. Reset action is thus provided and reset rate can be adjusted by means of variable resistor R_2. The proportional band of the control action can be adjusted at G and H.

Figure 10–26 shows the arrangement of the two lead networks, one in the forward path for rate action, the other as negative feedback for re-

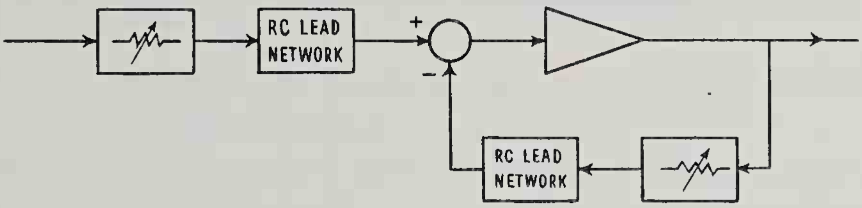


Figure 10–26. Schematic drawing of Leeds & Northrup controller.

set action. The two variable resistors express the alternate means for adjustment of proportional band. In this, as in practically all electric and electronic controllers, rate action is located before reset action. This reduces overshooting in those cases where the controlled variable is initially outside the proportional band as has been previously discussed.

Electric and electronic controllers, as described so far, use a position feedback obtained from the final control element. This is different from the conventional pneumatic controller which feeds back its own output signal. This output signal is then transmitted to the final control element.

The electronic controllers described in the following are in this and various other respects a close equivalent of the pneumatic controller. Figure 10–27 shows the principle of the Manning, Maxwell and Moore controller. The system is based entirely on d.c. signals. This refers to the output of the measuring element as well as to the output of the controller. The input to the controller is obtained from the primary feedback transmitted from the measuring means and a set point mechanism, the principle of which is illustrated in Figure 10–28. The primary feedback

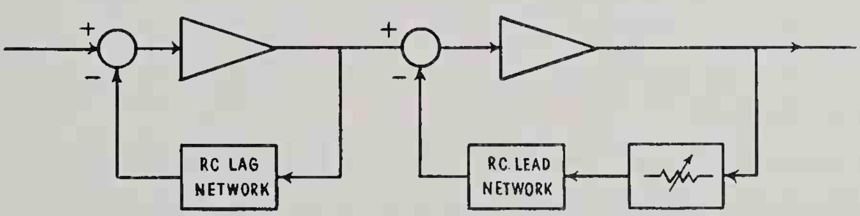


Figure 10–27. Schematic drawing of Manning, Maxwell and Moore controller.

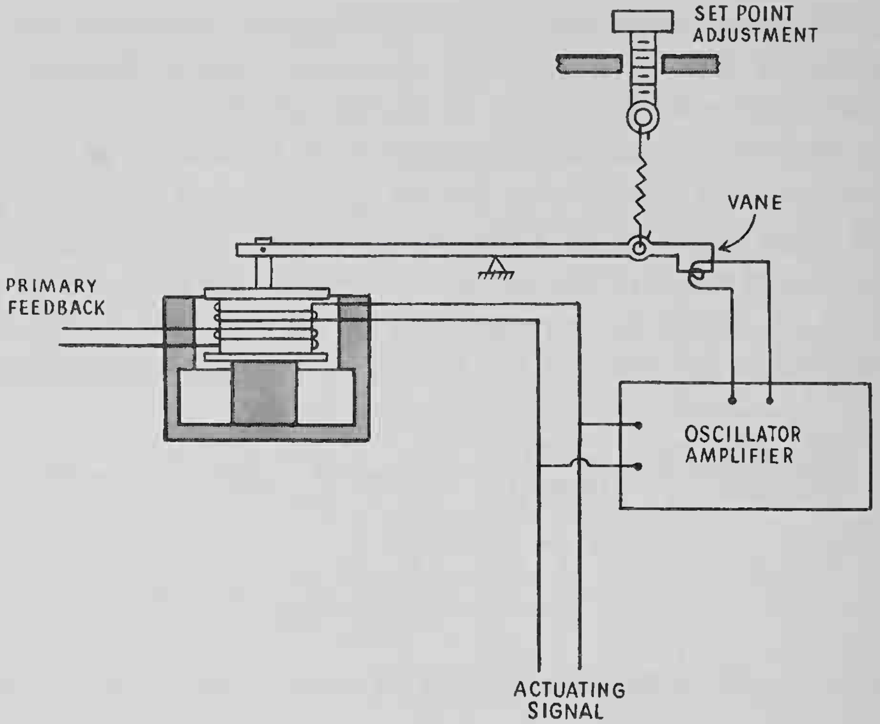


Figure 10–28. Principle of set point arrangement.

is applied to a moving coil. Suppose this signal increases, resulting in repelling the coil. This moves the beam, to which the coil is fastened, in a clockwise direction, and the vane will enter between the inductance coils of an oscillator amplifier. The d.c. output of the amplifier, which is the actuating signal, is thereby increased. The actuating signal is fed back to a second coil winding, which is wound so that an increase of current produces increased attraction and tends to move the beam counter-clockwise. The balance of the beam is determined by

(Actuating signal) + (Set point spring) = Primary feedback.

Hence, by changing the spring tension, the level of the actuating signal is raised or lowered, and the set point is thereby adjusted.

The resulting output is applied to the controller, illustrated in Figure 10–27. The first amplifier provides rate action, which is obtained by a lag network in negative feedback. If rate action is not required, this part of the circuit can be removed. A second amplifier provides the necessary voltage and power amplification. Proportional band is adjusted by a variable resistor in the negative feedback. A lead network provides the desired reset action.

The Swartwout controller requires a different approach, because the output signal of the measuring means is an a.c. signal derived from a differential transformer. This signal is matched with a second a.c. signal

which is manually adjustable and represents the set point. The resulting actuating signal is applied to the controller. The controller is illustrated in Figure 10–29. The first amplifier provides proportional band adjustment in its feedback. From there the signal must be converted into a d.c.

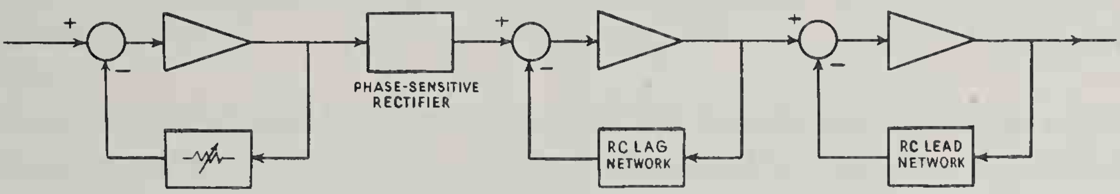


Figure 10–29. Schematic drawing of Swartwout controller.

signal in order to make *RC* lead and lag networks applicable. This is done in a phase-sensitive rectifier.

The phase-sensitive rectifier converts the signal which originated from a differential transformer. In doing so, it must be able to distinguish in which direction from the center the core of the differential transformer is displaced. The output in the midposition is zero and increases by equal increments when displaced from this position. The phase relationship to the primary is such that, assuming it is zero in one direction, it will be 180 degrees in the other direction. The phase sensitive rectifier responds to the phase shift by a shift of polarity in the direct current. It is then possible to raise the reference level by a summation circuit. For example, adding a constant voltage of 3 volts to a signal that changes between -2 and $+2$ volts, results in a signal changing from 1 to 5 volts.

Reset and rate action are provided by networks which in their principle are similar to those described before. The controller output is a d.c. signal.

Interaction

Practically all controllers—pneumatic, electric or electronic—exhibit some interaction between the adjustments of proportional band, reset rate, and rate time.

This means that a difference will generally exist between *nominal* settings of proportional band, reset rate and rate time and their *effective* values, depending on their relative adjustments. The change in effective proportional band* may be large enough to make the control system unstable when changing the reset rate or the rate time. The product of reset rate, in repeats per minute, and rate time, in minutes, should always be 0.25 or smaller. Otherwise instability may result.

* Young, A. J., "An Introduction to Process Control System Design," Instruments Publishing Co., Pittsburgh, Pa., (1955).

11. FINAL CONTROL ELEMENTS

The most common final control element is the sliding stem valve with pneumatic actuator as illustrated in Figure 11–1. The term actuator is used here in preference to "motor operator" which is recommended by the Automatic Control Terminology of the American Society of Mechanical Engineers.

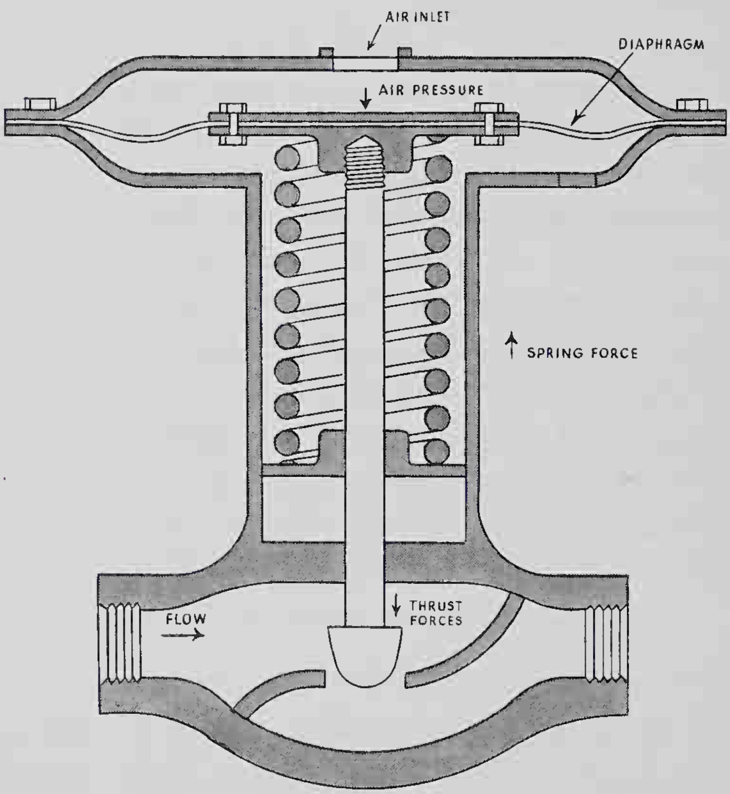


Figure 11–1. Diagram of sliding stem valve with pneumatic actuator.

The output signal of a pneumatic controller is applied to the top of the diaphragm. The range of the signal in standard applications is 3 to 15 psi. As the air pressure exceeds 3 psi the valve tends to move downward, compressing the spring. The force required to compress a spring is, in the ideal case, directly proportional to its deflection. In practice, however, a valve spring may deviate up to 5 per cent from linearity. Changes in temperature increase the nonlinear behavior even more.

Another source of nonlinearities is the change of effective area in the diaphragm described previously.

The relatively minor effect of nonlinearities in the final control element has been stressed before. It will be shown in the following pages that the action of the valve in the control loop includes a number of other nonlinear responses, all of which have nevertheless little if any effect.

At this point, it may only be pointed out that the nonlinearity, e.g., of the controller output signal, can go so far that it nullifies even the characterization of the control valve, and that, furthermore, particularly in flow and temperature control systems, the primary feedback signal is frequently far from linear.

Forces in a Spring-opposed Diaphragm Valve

The force exerted by the air pressure on the diaphragm compresses the spring until the force stored in the spring is equal to the sum of the forces from the air pressure, the friction and the thrust.

The friction is caused by the bearings for the valve stem, while the process fluid and its pressure drop across the valve port result in forces on the valve plug that act largely as vertical thrusts. The spring force must be large in order to minimize the effect of friction. The area of the diaphragm must be correspondingly large to be able to compress the spring.

Suppose the diaphragm area is 140 sq in. The minimum air pressure of 3 psi is to produce a force of 420 lb and the maximum pressure of 15 psi results in a force of 2100 lb. In the case of a 2-in. stroke, a spring constant of $(2100 - 420)/2 = 840$ lb/in. is required. The spring has to be precompressed by ½ in. in order to provide the initial force of 420 lb.

If the thrust is 280 lb, the pressure must increase to 5 psi before the valve begins to move. Similarly, if the valve is at midstroke and the thrust changes by 100 lb, the valve position will change by 0.12 in., or 6 per cent of its full stroke. This illustrates that a spring-opposed diaphragm valve should only be used where comparatively small thrusts are involved.

Friction acts in a similar manner. For example, a 20 lb breakloose force needs an increase of 0.143 lb in controller output signal, i.e., more than 1 per cent, before the control valve begins to move. This is equivalent to a dead zone of 2 per cent which may quite effectively contribute to the instability of a control system.

Stroke Time of the Spring-opposed Diaphragm Valve

The time needed to stroke a valve from fully open to fully closed depends largely on the effective area of the diaphragm. This makes the

smaller valves—which have smaller diaphragms—the faster ones. For a pressure change from 3 to 15 psig, typical stroke times in round figures are about as follows:

¾″ stroke valves—2 sec/in.
1½″ stroke valves—3 sec/in.
2″ stroke valves—5 sec/in.

These values, as any others based on a full-range pressure change, furnish only comparative data about the response in an actual control loop. In the first place, signal pressures change rarely over the complete range from 3 to 15 psi. In the second place, the effect of the valve action on the process decreases the signal from the controller and correspondingly slows down the speed. The speed of a given valve depends essentially on the port area of the pilot valve in the controller—disregarding time lags in the transmitting circuit.

Valve Positioners

The valve positioner is a device which compares the actual position of the valve with the controller output signal. As such it eliminates most of the shortcomings that are due to the above described method of using a spring force as an expression of the valve position. The valve positioner is generally an auxiliary device attached to the conventional spring-opposed diaphragm valves. In itself it may represent either a force-balanced or position-balanced mechanism.

Figure 11–2 is a diagram of a position-balanced valve positioner. The controller output signal is applied to a bellows which is spring opposed and assumes a position proportional to the signal. It moves a lever using point *A* as a fulcrum, which operates a pilot valve that admits secondary air to the valve diaphragm. The valve position is fed back by a linkage to the lever, turning it about point *B*. This repositions the pilot to a point where motion of the valve ceases. The valve has assumed a new position in proportion to the controller output signal applied to the bellows. Lever *C* can be adjusted to change the proportion which *A* moves with respect to valve motion. This alters the feedback and determines the valve stroke per unit change of controller output signal.

Figure 11–3 illustrates a force-balanced valve positioner. The lever is displaced against the force of the feedback spring. For example, the controller output signal increases, moving the bellows against the force of the feedback spring (neglecting the spring rate of the bellows). The pilot opens, admitting secondary air to the top of the diaphragm, moving the valve downward. This stretches the feedback spring, increasing its force and positioning the pilot and returning it to its initial condition,

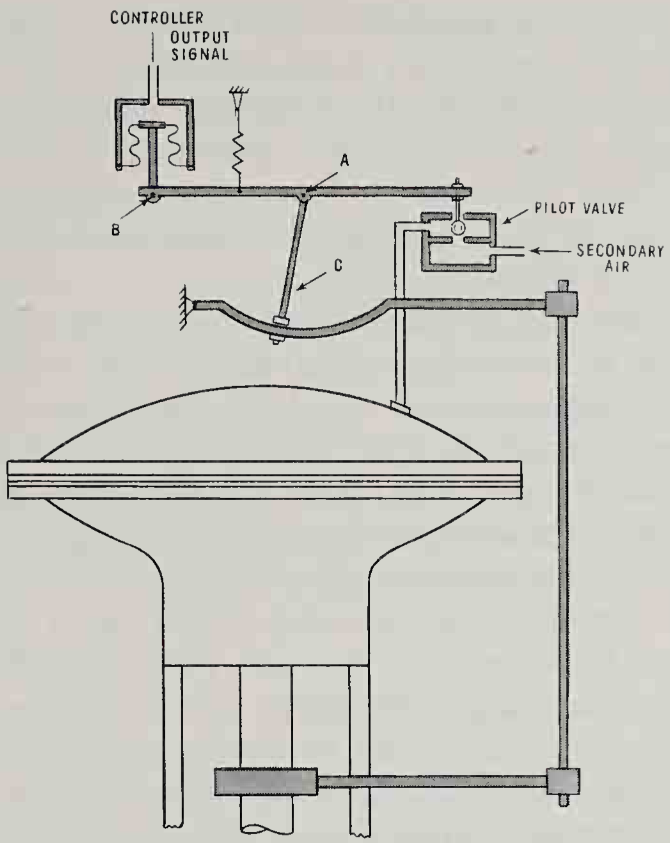


Figure 11–2. Position-balanced valve positioner.

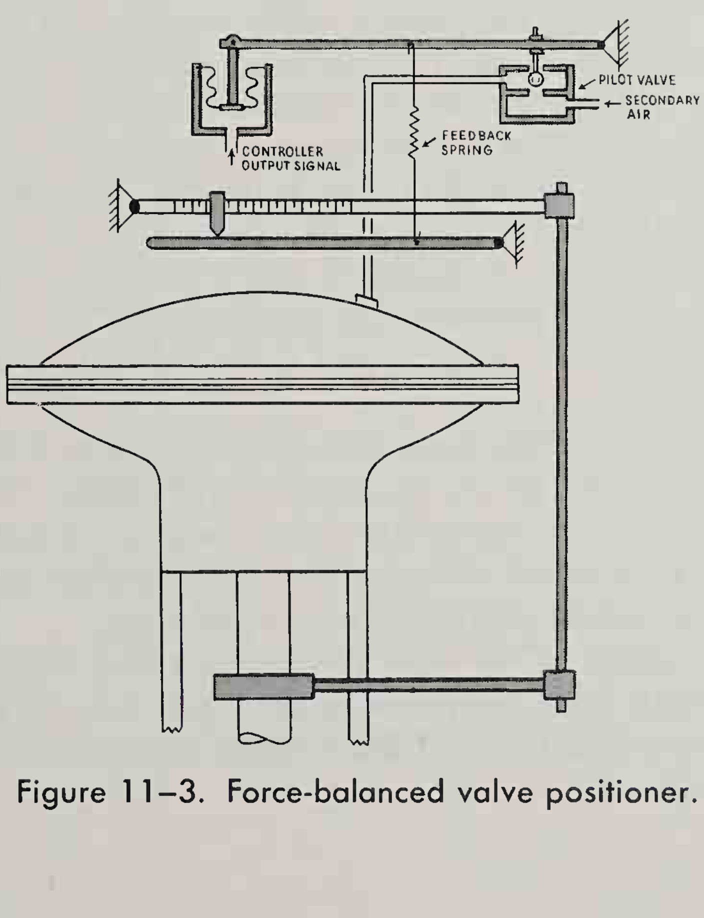


Figure 11–3. Force-balanced valve positioner.

at which point the valve settles in its new position. The stroke adjustment follows the principle of the position-balanced valve positioner. It is obtained by changing the feedback ratio by means of the lever ratio.

The valve positioner reduces the air volume involved in the controller output signal system, since the air space above the diaphragm is filled with secondary air. Within the limitations described below, this results in faster response of the valve.

A valve positioner will, in general, not increase the response speed of a valve if a change of the controller output signal from 3 to 15 psi is considered, unless the secondary air supply is more than 15 psi. This is because the pilot port area in the controller as well as in the positioner are of the same order of magnitude. However, for minor changes of controller output signal, the amplification of the positioner takes effect and produces faster response than can otherwise be obtained.

The positioner amplifies the controller output signal by a certain gain factor. This gain, however, does not affect the control loop as such. The control valve with the valve positioner represents a minor loop within the major control loop. This minor loop is composed of the controller output signal, the pilot valve, the valve stem motion, and the feedback to the controller output signal.

The gain of the positioner is limited by the pressure of the secondary air supply. A typical valve positioner may require a change of 0.07 psi in the controller output signal to move the pilot valve from fully closed to fully open. If the control valve is closed at 3 psi and the secondary air supply is 15 psi, then an increase of the controller output signal from 3 to 3.07 psi or more will produce an initial pressure drop of 12 psi across the pilot. This results in correspondingly high rate of flow to the diaphragm top, which decreases gradually as the new position is reached. On the other hand, suppose the valve carries at a given position 14 psi on the diaphragm. A change in controller output signal of 0.07 psi in controller output pressure will now result in an initial pressure drop of only 1 psi across the pilot and the valve motion will be correspondingly slower.

The gain and the resulting time constant of a control valve with valve positioner depends therefore on the position of the valve. This is particularly noticeable when the secondary supply pressure is not higher than 15 psi. This behavior of the valve positioner can explain instability of a control system at certain controller output pressures.

The main advantage of the valve positioner is that in general it has a comparatively high gain, and much larger forces can therefore be produced to overcome sticking and friction of the control valve than is possible without positioner.

Due to the high gain of the valve positioner, it reduces considerably the tendency of the valve to change its position with variations in thrust. An increase in thrust, for example, will generally not be able to displace the position by more than one per cent without increasing the diaphragm pressure to the maximum.

Fast changes in thrust may, however, change the valve position because of the time lag in the circuit. In general, these position changes are not harmful, or at least less harmful than the pressure changes in the process fluid which produce the variations in thrust. Accurate control can hardly be obtained where sudden pressure changes in the process fluid occur, and the changing valve position may become the visible expression of a more basic trouble in the control system. If the thrust changes are slow they may cause some response in the valve position but without affecting control in a closed system.

Valve Characteristics

Nonlinearities may exceed acceptable limits. This is particularly the case in the relation between valve port area and flow through the valve as well as in the process response itself. To reduce these nonlinearities, valve plugs of different shapes, as shown in Figure 11–4, are available. These plugs have nonlinear characteristics in their relation between valve lift and valve port area. This relation is called the *valve characteristic.*

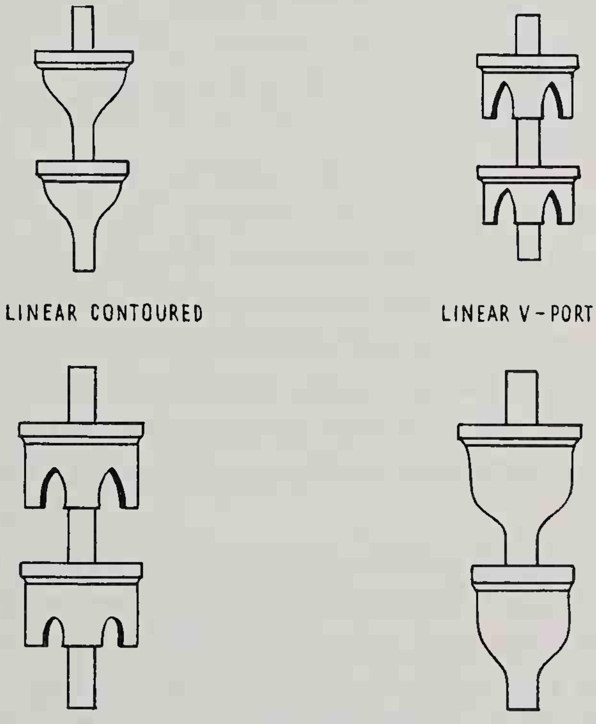


Figure 11–4. Typical valve plugs.

Figure 11–5 illustrates the open loop of a temperature control system. The input to the controller is the measurement signal without conversion, i.e., the input changes 1°F per 1°F change in the controlled variable.

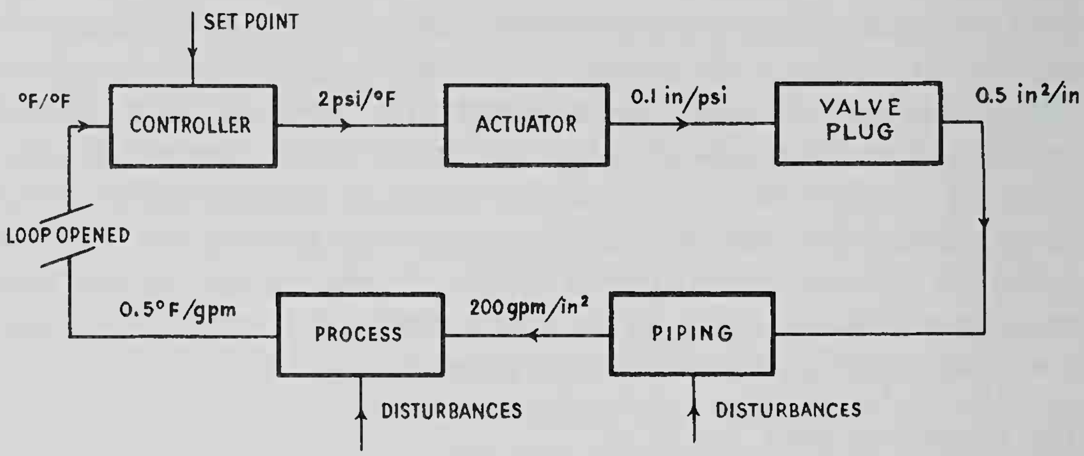


Figure 11–5. Temperature control system.

The controller output signal changes 2 psi/°F change in input. This signal is converted by the valve actuator into a valve stroke with a gain of 0.1 in./psi. The valve port area changes 0.5 sq in. for each inch in valve travel, assuming a linear valve characteristic. The flow through the pipe line varies at the rate of 200 gpm/sq in. port area, and the temperature responds to flow changes with 0.5°F/gpm. The total loop gain is expressed by °F/°F $\times$ 2 psi/°F $\times$ 0.1 in./psi $\times$ 0.5 in.^{2}/in. $\times$ 200 gpm/in.^{2} $\times$ 0.5°F/gpm = 10

This gain should be maintained—ideally—for any set point, for any pressure drop, and for any load conditions. The fact is that the gain in the piping, for example, is not constant. Where this occurs, compensation may be possible by characterization of the valve plug. For example, if the gain in the pipe line changes to 100 gpm/in.^{2} while simultaneously the gain in the plug increases to 1 in.^{2}/in., the overall gain would remain the same and the dynamic performance of the system would not be affected. The ideal valve characteristic maintains a constant gain in the control loop.

Figure 11–6 shows valve characteristics corresponding to the plugs illustrated in Figure 11–4. Valve characteristics are not tailor-made. Improvements can be expected, although perfectly constant gain throughout the system and operation is an unattainable ideal.

The flow rate through a valve is expressed by the equation

$$Q = k_1 A \sqrt{p} \qquad (11\text{–}1)$$

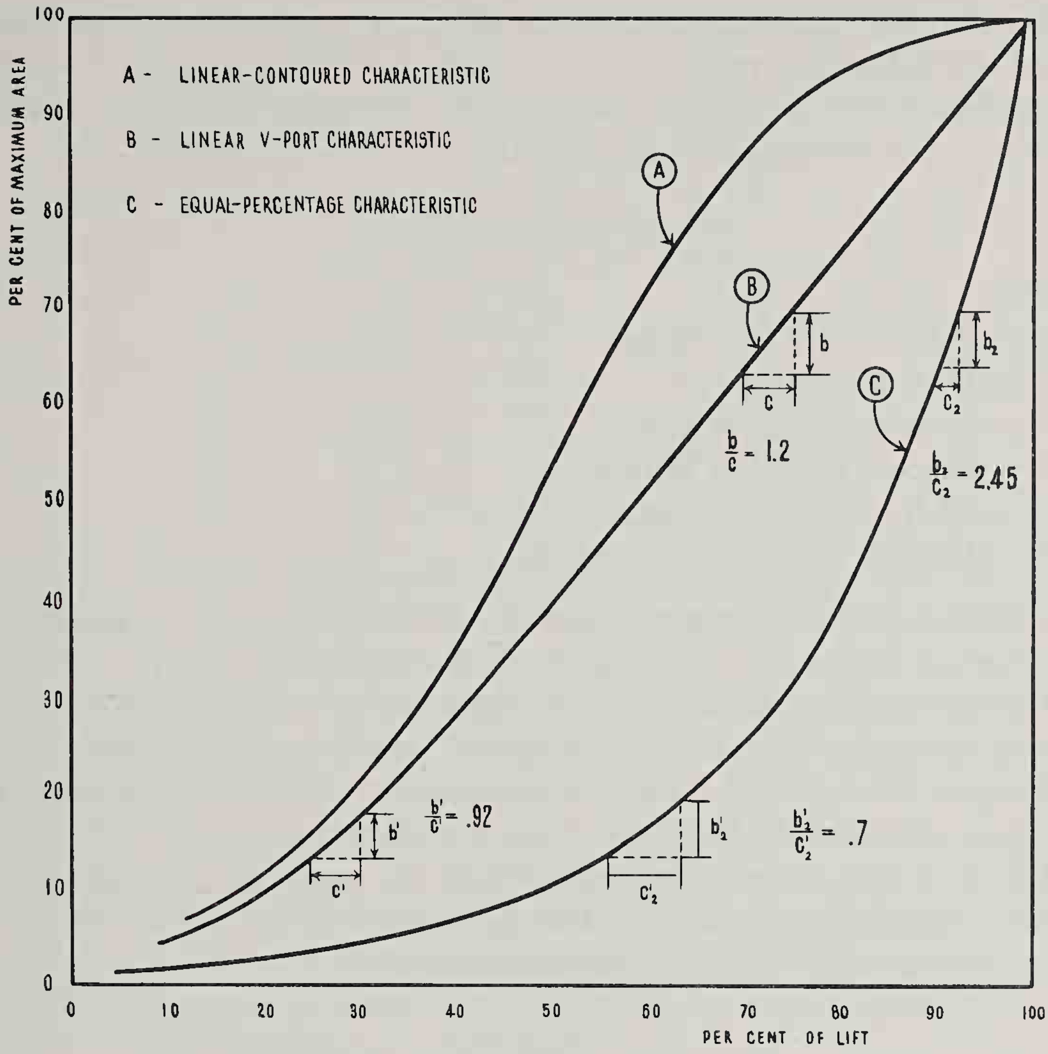


Figure 11–6. Valve characteristics.

where Q is the flow rate, k_1 a constant, A the port area, and p is the pressure drop across the valve port. As long as p remains constant, the relation between flow rate and port area is linear. There are, however, two additional pressure drops changing with the flow rate and hence affecting p. These are the pressure drop in the valve body and in the pipeline. Generally, the drop in the valve body can be neglected, which will be done in this discussion.

Considering a total available system pressure P and a pressure loss in the pipeline of p_1, the pressure drop across the valve is then given by

$$p = P - p_1$$

and hence equation (11–1) may be written

$$Q = k_1 A \sqrt{P - p_1} \qquad (11\text{–}2)$$

Obviously, the smaller can p_1 be made in proportion to P, the more linear will be the operation.

Pipeline loss is generally calculated by means of Fanning's equation which, in one practical version, reads

$$p_1 = 0.0066 \frac{sL}{D^5} fB^2$$

where

p_1 is the pressure loss in psi,
s is the specific gravity,
L is the pipe length in feet,
D is the pipe diameter in inch,
f is the friction factor, and
B is the flow rate in barrels per hour.

The friction factor is obtained from a suitable table, such as given in the *Mechanical Engineers' Handbook* by L. S. Marks.

The rate of flow for various port areas can be calculated as in the following example. The fluid is crude oil of 0.9 specific gravity which is pumped through 300 ft of 2-in. wrought iron pipe. Normal maximum pumping rate is 400 barrels/hr. Total pressure available is 250 psi.

The first step consists in determining the pipeline pressure drop at various pumping rates. This is done by means of Fanning's formula and the results are shown in the following table:

TABLE 1.

Bbl/hr	p_1
40	1.6
120	12.2
200	32.5
300	68.0
400	118.0

Using conventional methods, a valve size is now chosen, which in this case shall be 1.5 in. This corresponds to an approximate port area of 1.75 sq in. It may be assumed that the normal maximum flow of 400 barrels per hour corresponds to a port area of 80 per cent of maximum. Inserting these values in equation (11–2) and solving for k_1 gives

$$k_1 = \frac{400}{(1.75 \times 0.8)\sqrt{250 - 118}} = 25$$

This value of k_1 can now be used to solve by means of equation (11–2)

for the valve port area at the different values of flow and pipe line pressure drop as given in Table 1. The results are listed in Table 2.

TABLE 2.

Bbl/hr	p_1	A	% of max. area	% of max. flow
40	1.6	0.1	6	9
120	12.2	0.3	17	27
200	32.5	0.5	29	45
300	68.0	0.9	52	67
400	118.0	1.4	80	90
445	145.0	1.75	100	100

The last row of values has been obtained by inserting a few trial-and-error values in the calculations for Table 1 to obtain the flow which corresponds to the maximum port area of the valve.

The results of Table 2 are graphically represented in Figure 11–7. The gain of change in flow rate *vs* change in port area at any point can be easily determined as shown in the illustration. At about 15 per cent of maximum port area the gain is 1.64. At about 68 per cent the gain has decreased to 0.7. In Figure 11–6, the gain for these same port openings has been shown on the linear-contoured and the equal-percentage characteristics. Thus two graphs are available: one for the pipeline characteristics (Figure 11–7), the other for the valve characteristic (Figure 11–6). It remains to correlate these data.

The gain of flow rate over valve lift is the product of

$$\frac{\text{Port Area}}{\text{Lift}} \times \frac{\text{Flow}}{\text{Area}} = \frac{b}{c} \times \frac{a}{b}$$

The symbols a, b and c refer to Figures 11–6 and 11–7 and may be replaced by a', b', c', b'_2, c'_2, or b_2, c_2, whichever of the small triangles is referred to in the graphs. Table 3 shows the various gains:

TABLE 3.

Port Opening	A Linear V-Port	B Equal Percentage	C Pipeline	Combined Gain	
				A × C	B × C
15%	.92	.7	1.64	1.51	1.14
68%	1.2	2.45	0.7	0.84	1.71

In this example, it appears that the equal-percentage characteristic would give somewhat better results since it changes its gain by 1.71:1.14 =

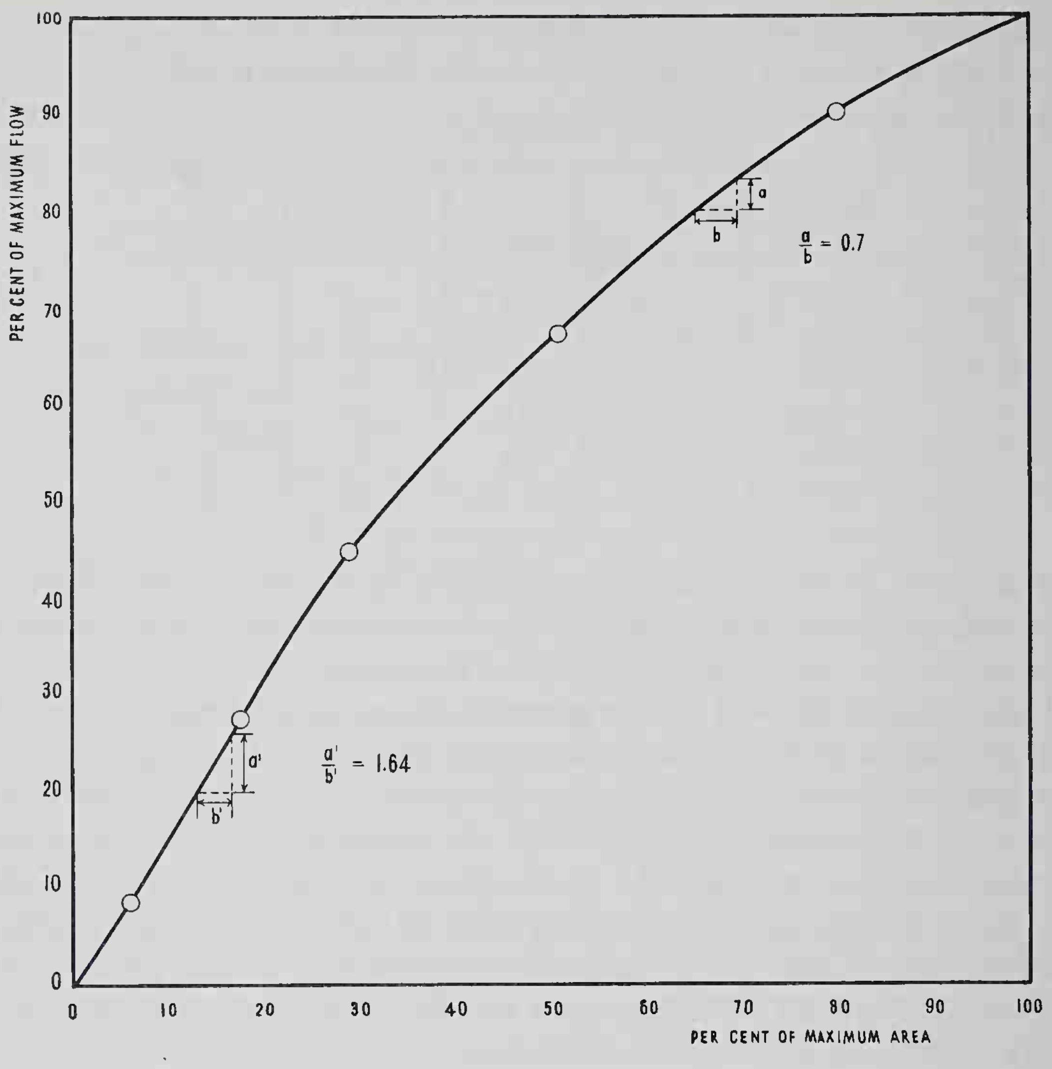


Figure 11–7. Flow-area characteristic of pipeline.

1.5, while the linear V-port characteristic changes the combined gain by 1.51:0.84 = 1.8. However, in the first place, the difference is smaller than the uncertainties which are inherent in the calculations; in the second place, comparison would be necessary between more than two points. In fact, these points must be carefully chosen to be in line with the most probable operating ranges. Evaluation of gain changes in the process will further influence the choice of the valve characteristic. This systematic approach, however, is only a further application of the method outlined here. Important are the relatively high nonlinearities which have to be accepted.

Rangeability

Characterized valves are not able to operate over the full range. As the valve approaches its closed position, the valve characteristic cannot even approximately be maintained, and a valve operating in this range

would give erratic results in a control system. This limitation is expressed by the *rangeability factor.* If the valve can be used between 10 and 100 per cent of its stem travel, the rangeability is $^{100}/_{10} = 10$. If it is limited to a range between 20 and 100, it is $^{100}/_{20} = 5$. A good practical figure is an average rangeability of 8, approaching 15 on larger valves, but decreasing to as much as 5 on small valves.

In determining the rangeability required for an actual case, it is necessary to take into account pressure drop across the valve as well as flow. Calling the maximum flow Q_1, and the corresponding valve pressure drop p_1, and calling the minimum flow Q_2 and the corresponding valve pressure drop p_2, the required rangeability R is given by

$$R = \frac{Q_1}{Q_2}\sqrt{\frac{p_2}{p_1}}$$

For example, if the maximum flow were 360 gpm at 12 psi and the minimum flow 60 gpm at 27 psi, the required rangeability would be

$$R = \frac{360}{60}\sqrt{\frac{27}{12}} = 9$$

Process Gain

It is generally assumed that the process gain is constant under altering load conditions. Closer analysis of a process may show that this is not so and a valve characteristic may be chosen for partial compensation of these gain variations.

Changes of set point will affect the gain of almost any process. A tuning of the controller for one set point may cause instability or excessively slow corrective action at another set point. It becomes difficult, if not impossible, to consider all the various effects and combinations in selecting a valve characteristic. However, it is never lost time to consider certain possible combinations of (a) valve characteristic, (b) flow through the valve, (c) load, and (d) set point, and to determine the gain under these conditions. Conditions may be determined beforehand which have the highest gain and are most likely to produce instability under actual operation.

Conclusions on Valve Action in Control Loop

The question may well be raised how it is possible that satisfactory control can be obtained at all under conditions where disturbing nonlinearities may be reduced but not eliminated. To answer this, static and dynamic conditions must be considered separately. The nonlinearities may produce a static inaccuracy, and hence a deviation of the controlled

variable from the set point. In floating controllers and those that include reset action, these deviations are automatically corrected by the controller action. In proportional controllers, the offset resulting from load changes is generally greater than similar consequences of the nonlinearities. Therefore, in either case, considerable nonlinearity can be tolerated without having effect on the control system.

Dynamically, the nonlinearities require a tuning of the controller that assures stability under the conditions of highest gain. In practical operation, controllers are usually adjusted with enough stability margin to assure stable operation even when conditions are occasionally much closer to instability than originally expected. However, it is here where nonlinearities may occasionally cause trouble, since the controller is adjusted for specific load conditions or a certain set point, and changes in either may produce instability and require changes in the controller adjustments.

Other Control Valves and Actuators

So far only the spring-opposed diaphragm valve has been considered. Most of the features are similar to other types like butterfly or Saunders Patent valves, and to other actuators or combinations like the electropneumatic relay or the electrohydraulic valve actuator. From the viewpoint of control systems, some particulars of the butterfly valves and the final control elements in electronic control systems are of special interest.

Butterfly Control Valves. Butterfly valves are particularly suited for controlling large flows, especially at low pressures. They are also used to advantage in lines carrying considerable amounts of suspended matter, which would cause excess clogging of plug-and-seal types of valves. They usually operate through about 60 degrees of their rotary movement. These valves produce practically no pressure drop when they are fully open. A butterfly valve is shown in Figure 11–8.

The unbalanced forces across the discs of butterfly valves, which the actuator has to overcome, are comparatively high. The tendency of these forces is to close the valve. Their magnitude is proportional to the pressure drop across the valve and the cube of the disc diameter, and depends on the disc position.*

The unbalanced forces push the disc against its bearings and friction results. This friction must be overcome by the actuator. In closing the

* Dally, Charles A., "Butterfly Control Valves," *Instruments* (Dec. 1952).

valve, the actuator is assisted by the unbalanced forces, but in opening it is opposed by the friction plus the unbalanced forces themselves. The torque required to position a large butterfly valve, e.g., 24 in., may amount to 6000 foot-pounds and more under normal operating conditions.

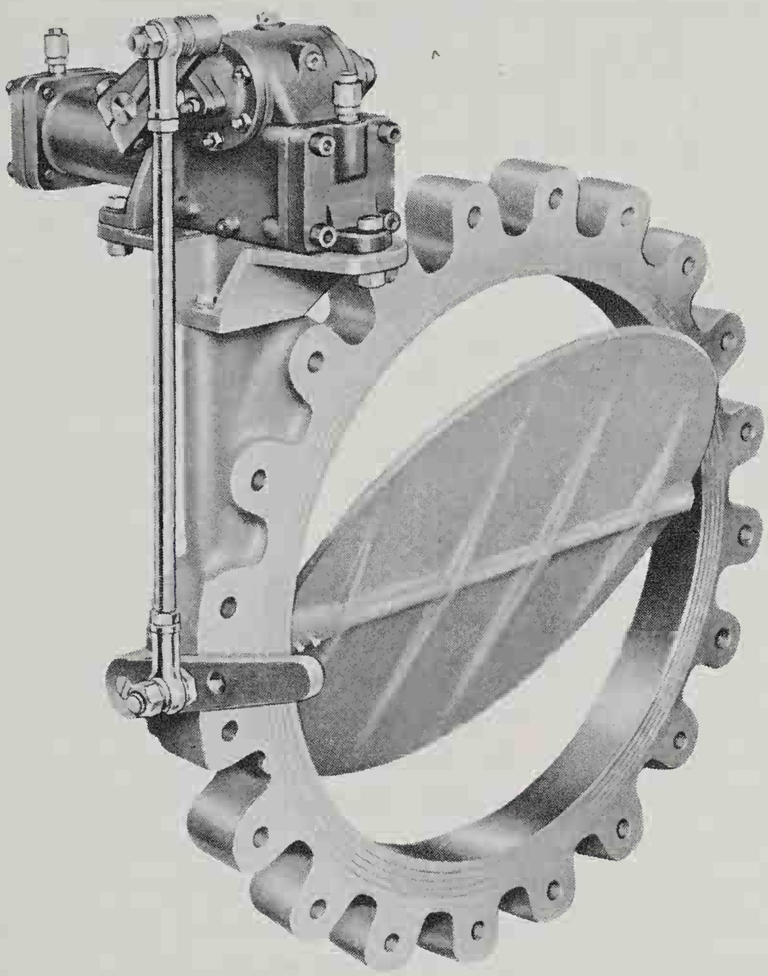


Figure 11–8. Butterfly valve.

Such large and changing torques will have an influence on the control system to be chosen. Pneumatic systems are frequently no longer adequate and hydraulic systems will be preferred.

Electropneumatic Relays. The electropneumatic relay converts the output signal of an electronic controller into a pneumatic signal which in turn is applied to the conventional spring-opposed diaphragm valve with or without positioner. Figure 11–9 illustrates the principle as used in the Swartwout Power Relay. The controller output signal, i.e., the input

signal into the relay, is applied to a moving coil which is mounted on a lever. On the right the lever is fastened to a feedback diaphragm. When the electrical input signal increases, the resulting increase of force in the moving coil tends to move the lever away from the magnet. This depresses

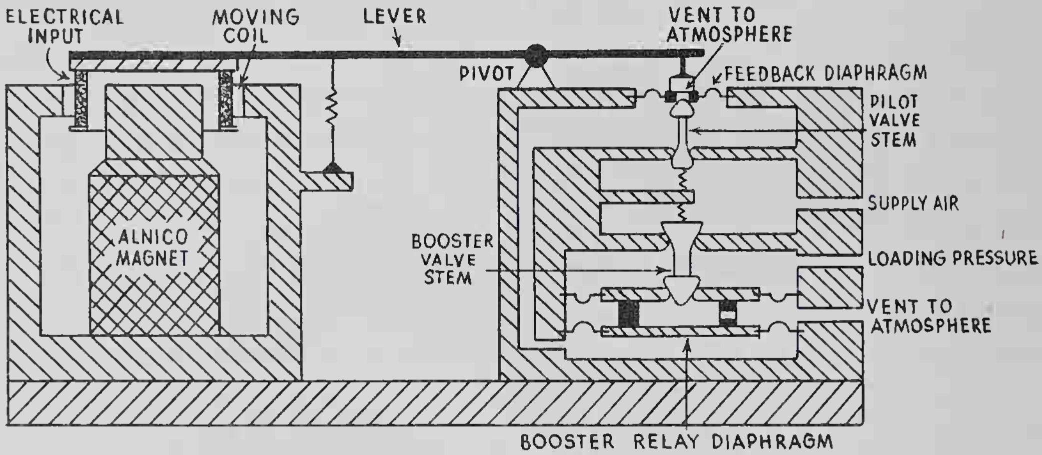


Figure 11–9. Electropneumatic relay—Swartwout principle.

the feedback diaphragm. Its downward motion displaces the pilot valve stem, opening the lower port of this valve. The result is that the air pressure under the feedback diaphragm as well as under the booster relay diaphragm increases. The pilot valve remains open to the air supply until the pressure under the feedback diaphragm has increased sufficiently to balance the force due to the electrical input signal.

The increased air pressure under the booster relay diaphragm moves the booster valve stem upward and opens the upper port of this valve. Supply air then flows through the port, increasing the loading pressure and the pressure on the upper side of the booster relay diaphragm. The upper port of the booster valve closes as soon as the pressures on both sides of the diaphragm are equal. This reestablishes balanced conditions, but at a loading pressure increased in proportion with the increased electrical signal.

The action is similar on decreasing electrical signals. In either case the loading pressure follows the changes in input signal in linear relation.

Electrohydraulic Actuators. A diagram of an electrohydraulic valve actuator is shown in Figure 11–10. The moving coil is fastened to a jet pipe. When the signal current increases, the coil is pulled into the magnetic field and the jet pipe rotates in counterclockwise motion. The result is that the piston moves downward together with the plug of the control valve to which it is linked. The feedback lever is connected to the piston stem

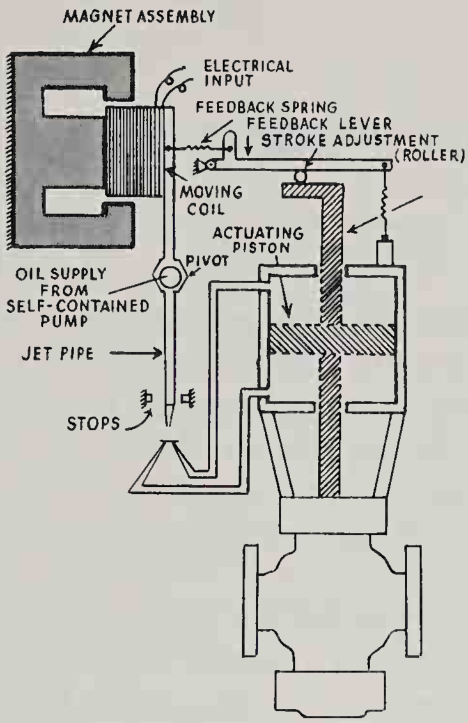


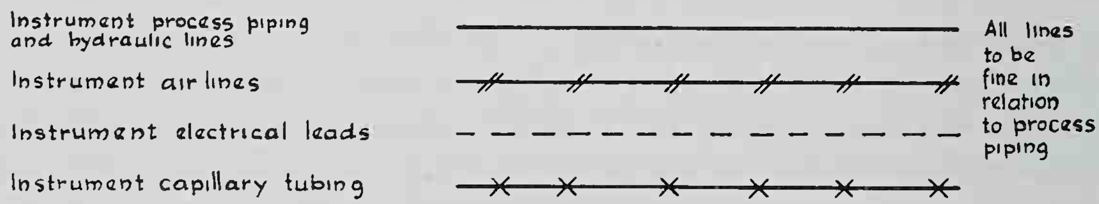
Figure 11–10. Schematic drawing of Askania electrohydraulic valve actuator.

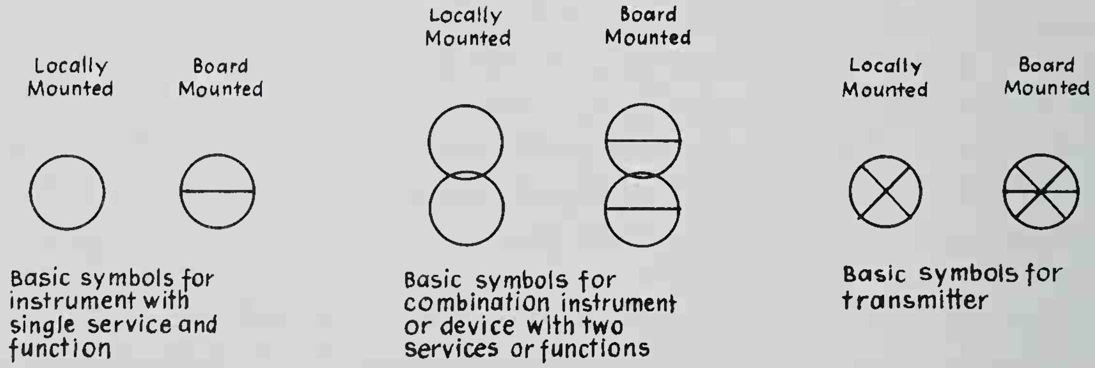
by means of a roller and a spring and follows its motion. This results in a stretching of the feedback spring as the piston moves toward a new position. As the force of the feedback increases and approaches the opposing force of the moving coil, the jet pipe returns to its neutral position, and once this position is attained, the piston stops. The position of the piston is proportional to the magnitude of the signal current.

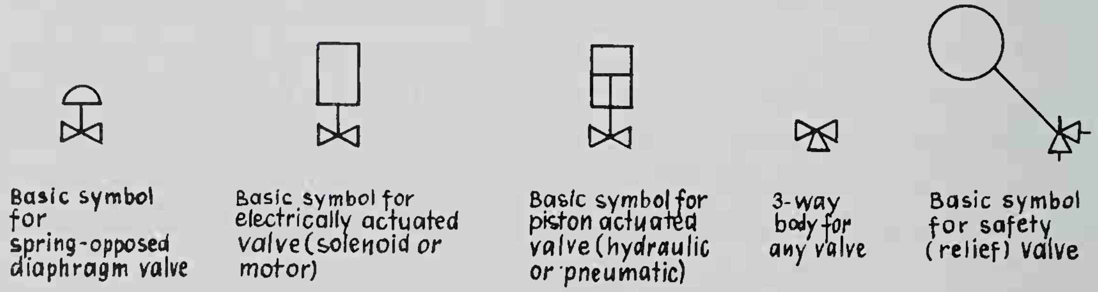
12. CONTROL SYSTEMS

It is one thing to consider a controller, e.g., a temperature controller, as an isolated phenomenon that is connected into a closed loop with some imaginary process unit, and another to be aware of the numerous details that influence the action—and hence the selection—of a control system within the totality of an industrial production arrangement. The difference is one between specifying control components and engineering a control system.

In illustrating control systems it is convenient to use a system of symbols and identifications which allows representation in the form of a flow plan. Such a system has been developed by the Instrument Society of America and is contained in their "Recommended Practices."







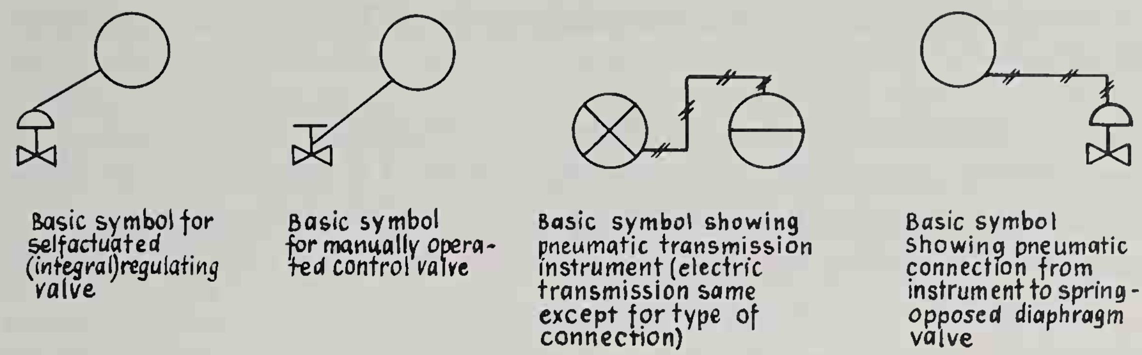


Figure 12–1. Basic instrumentation symbols.

Basic instrumentation symbols are shown in Figure 12–1. In order to identify control components and instruments in a flow plan the following letters and combinations are used.

Letters of Identification
Definition and Permissible Positions in Any Combination

LETTER	FIRST LETTER Process Variable or Actuation	SECOND LETTER Type Reading or Other Function	THIRD LETTER Additional Function
A	—	Alarm	Alarm
C	Conductivity	Control	Control
D	Density	—	—
E	—	Element (Primary)	—
F	Flow	—	—
G	—	Glass (No Measurement)	—
H	Hand (Actuated)	—	—
I	—	Indicating	—
L	Level	—	—
M	Moisture	—	—
P	Pressure	—	—
R	—	Recording (Recorder)	—
S	Speed	Safety	—
T	Temperature	—	—
V	Viscosity	—	Valve
W	Weight	Well	—

The following may be used optionally as a first letter for other process variables:

1. "A" may be used to cover all types of analyzing instruments.
2. Readily recognized, self-defining chemical symbols such as CO_2, O_2, etc., may be used for these specific analysis instruments.
3. The self-defining symbol "pH" may be used for hydrogen ion concentration.

Furthermore, it is permissible to insert a lower case "r" after "F" to designate Flow ratio. Likewise, lower case "d" may be inserted after "T" or "P" to designate Temperature difference or Pressure difference.

If the flow contains several units of equal letter identification, it becomes necessary to supplement the general identification by a numerical system, to establish its specific identity. Thus, the identification TRC–1, refers to a temperature recording controller and to item No. 1.

Figure 12–2 applies this system for the typical instrumentation of a fractionating column. The feed to the fractionating column is maintained constant at a certain rate which is established by the set point of the flow recording controller (FRC–2). This set point is automatically adjusted by the level recording controller (LRC–1). The level control maintains not a fixed but an average level. This arrangement combines average and cascade control as will be described further below. Its purpose is to reduce fluctuations in the feed flow rate as much as possible.

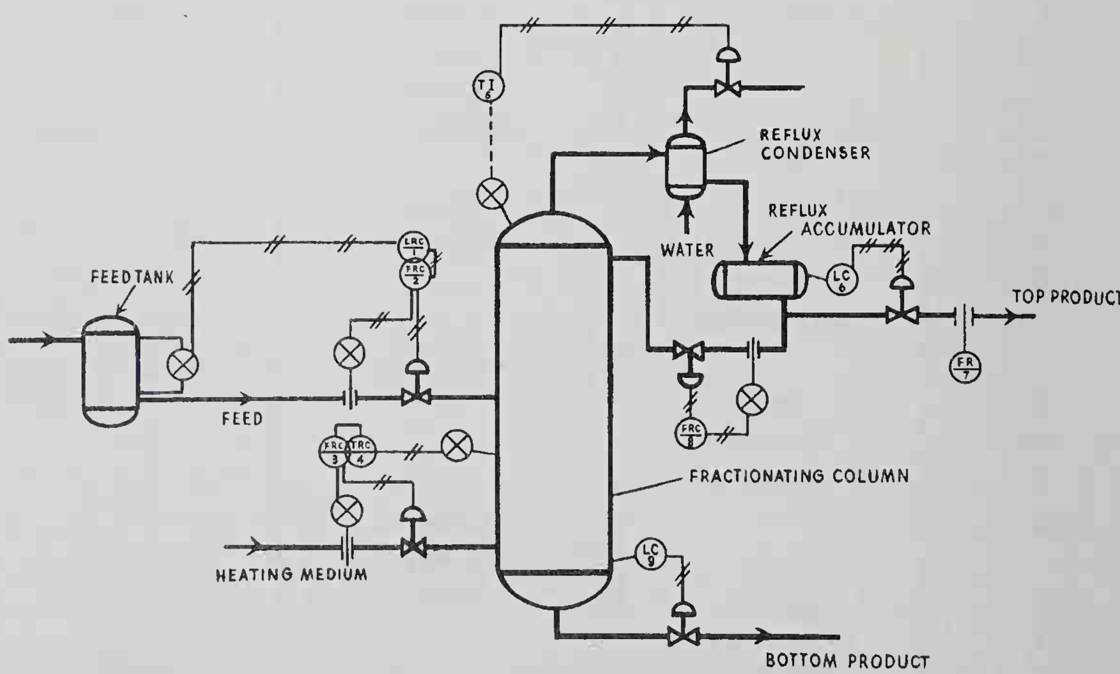


Figure 12–2. Typical instrumentation of fractionating column.

The situation is somewhat similar in the control of the flow rate of the heating medium. The flow rate has to be kept constant even though upstream or downstream pressures may change. On the other hand, temperature in the fractionating column must be kept constant. Hence the purpose of the cascade control in this case is to change the set point of the flow recording controller (FRC–3) by means of the temperature recording controller (TRC–4). The remaining control loops in this case are single-loop controls and need no further description as such.

Averaging Control

In the feed tank of Figure 12–2, it is quite permissible for the level to fluctuate between certain limits, but it is of greatest importance that the flow to the fractionating column does not change abruptly. A control system of this kind, purposely permitting variations of the controlled variable which are larger than required by the system dynamics, is called an *averaging control system.*

Suppose the maximum permissible rate of change in flow is 3.75 gpm/min. For a control valve with linear characteristic and a flow capacity of 37.5 gpm at the prevailing pressure, the minimum time for full stroke should be 10 min. This can be effected with hydraulic controllers by selecting the corresponding flow rate of the actuating medium and the piston area. In pneumatic controllers, conditions are such that the timing of the final control element is not practical. The approach to the problem is then different and will be discussed further below.

Before the speed of the final control element is adjusted, it is necessary to ascertain that the rate at which the flow into the tank changes is always considerably below 3.75 gpm/min., or that it exceeds this rate only intermittently. The time during which the flow into the tank can exceed the maximum rate of change depends on the capacitance of the tank, the rate of flow, and the tolerance between maximum and minimum level. With feed tanks, the tolerance is generally so high that considerable amounts can be absorbed. Furthermore, limit controls may be provided to take safety measures in case these limits are exceeded.

Boiler drums are more critical in this respect. In their case the flow into the drum can be changed only gradually to avoid shocks. The level is also allowed to fluctuate but ± 2 in. is usually considered the limit. Furthermore, the flow rates may be considerably larger than those for the feed drum. The load may change rapidly, but careful study should be given to the possibility that the load might change faster than the feedwater flow, draining the boiler drum beyond the minimum level before the control valve in the feed line can follow.

It was stated that in pneumatic control a simple speed adjustment of

the final control element cannot be used. Hence, while for the hydraulic controller a proportional-speed floating controller could be used, proportional-position or a proportional plus reset action would be the choice if the final control element is pneumatically actuated. The point is then to select the widest possible proportional band. A method for determining the proportional band is described in the following paragraphs.

For a tank of 8 ft in diameter, i.e., an area of about 7200 sq in., and a permissible level variation of ± 20 in., the volume contained between minimum and maximum level is approximately 1250 gallons. Figure 12–3 illustrates this case. For a maximum flow of 250 gpm into the tank,

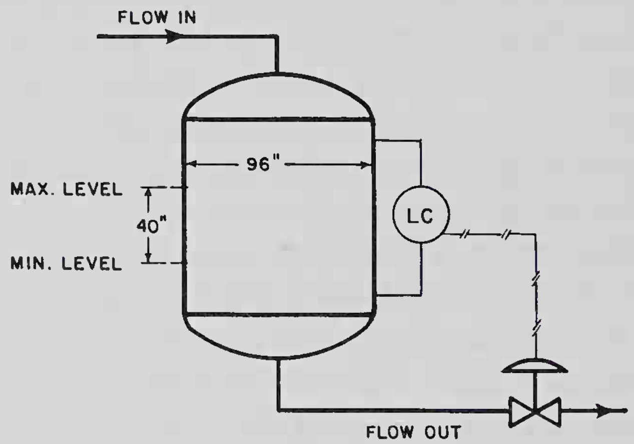


Figure 12–3. Averaging control system.

it is necessary that the valve reaches a position at maximum level at which 250 gpm are allowed to flow out of the tank.

A minimum flow of 50 gpm into the tank may exist, and hence at minimum level the valve should still be sufficiently open to allow an outflow of 50 gpm. If the valve is rated at maximum capacity of 300 gpm, this means that the controller output signal should be 5 psi at minimum level and 13 psi at maximum level. Since these levels are 40 in. apart, the proportional band of the controller is a 60 in. change of the controlled level, i.e., the flow rate changes by 5 gpm for every inch of change in tank level.

The proportional band is thus given by

$$P = \frac{V}{Q} L \tag{12–1}$$

where

P is the proportional band expressed in inches of controlled tank level,
V is the maximum flow capacity of the valve, in gpm,
L is the tolerance between maximum and minimum level, in inches, and
Q is the maximum sudden change of flow into the tank, in gpm.

The valve capacity depends on the level in the tank. The maximum level is the most critical. It is hardly conceivable that the maximum level coincides with the maximum flow change, but this condition may be assumed in order to be on the safe side. It is also assumed that the valve has linear characteristics. Such a valve could be employed in this system. If valve characteristics are changed, the fastest change in rate for a given displacement should be established.

The greatest and fastest change of flow into the tank would be a sudden increase of 200 gpm, from 50 to 250 gpm. Since the volume between maximum and minimum level is 1250 gallons, the level rises

$$\frac{200 \text{ gpm} \times 40 \text{ in.}}{1250 \text{ gal.}} = 6.4 \text{ in./min.}$$

Consequently, the valve would initially open at the corresponding rate of 5 gpm/in. rise in level which is 32 gpm/min.

This procedure can be expressed in the equation.

$$q/\text{min.} = \frac{QLV}{eP} \tag{12–2}$$

where

q/min. is the rate of change of flow obtained by valve action, in gpm/min.
Q is the maximum sudden change of flow into the tank, in gpm
e is the tank capacity between maximum and minimum level, in gal.

The other factors correspond with those of equation (12–1). Obviously, in either case any other consistent units may also be used.

This equation can be used in various ways. For example, the fastest permissible rate of change of flow is given as 16 gpm/min., while all other conditions are the same as before. In this case, q/min. = 16 and insertion of values into equation (12–2) allows solving for the proportional band P. The result is a proportional band of 120 in.

Since the equation contains the tank capacity between maximum and minimum level, it permits selection of the correct tank for a given set of conditions.

It has been assumed in the preceding that Q, the maximum sudden change of flow into the tank, is equivalent to a change from minimum to maximum flow into the tank. Actually such a sudden change may never happen, or if it does happen, it is permissible that limit switches open dump valves or take similar precautionary measures. Under such conditions, the value of Q may be chosen correspondingly smaller.

Certain modifications are possible in the assumptions contained in the foregoing. For example, the flow into the tank may reach its maxi-

mum only intermittently. In this case the Q in equation (12–1) is substituted by the factor M, which expresses the difference between the maximum and minimum flows that can be expected. Hence

$$P = \frac{V}{M} L \tag{12–3}$$

The proportional band P, becomes now correspondingly larger and its insertion in equation (12–2) results in either a smaller tank or a slower maximum rate of change of flow from the tank. It should be noted that the factor Q in equation (12–2) remains unaltered, since it is assumed that fluctuations through the maximum range may occur intermittently. If this is not the case, further corrections are possible.

When using equation (12–1), a proportional-position controller is to be used. The case of equation (12–3), however, requires proportional plus reset control. The reason is that offset decreases the absorption potential of the tank. Using a proportional-position controller without reset means considerable offset in view of the large proportional band used in averaging control. The control point may therefore drift considerably from its set point and be near one of the level limits when a sudden load change occurs. As long as the proportional band is adjusted according to equation (12–1), the level limits can hardly be exceeded. No advantage from reset action can be expected.

If the sensitivity of the controller is further decreased, in accordance with equation (12–3), then reset action is required to assure that the control point is always driven back to the set point, and that sufficient absorption capacity in the tank is available for sudden load changes. It is advisable that limit switches are provided in such arrangements to provide dumping or similar precautionary measures in case load changes are occasionally larger or more persistent than expected.

Cascade Control

Cascade control has two purposes. One is to control for two or more different kinds of load changes. The other is to improve the dynamics of a system by effectively reducing the dead time of the system.

Figure 12–4 shows the level control system of a feed tank, similar to that in Figure 12–2. Again the main purpose is to avoid too rapid changes in the feed flow. Averaging liquid level control is used. However, this protects only against sudden changes of flow into the feed tank. In this case another variable becomes important, namely, upstream and downstream pressure changes. The flow must be kept constant, no matter what the pressure conditions are. It still must change with level. In other words, two kinds of load changes must be controlled: pressure

and tank level. The solution is illustrated in both the flow diagram and the block diagram of Figure 12–4. A second control system is connected into the level control loop. The level controller "cascades" into the flow controller. The method is basically simple. All that is required is a flow

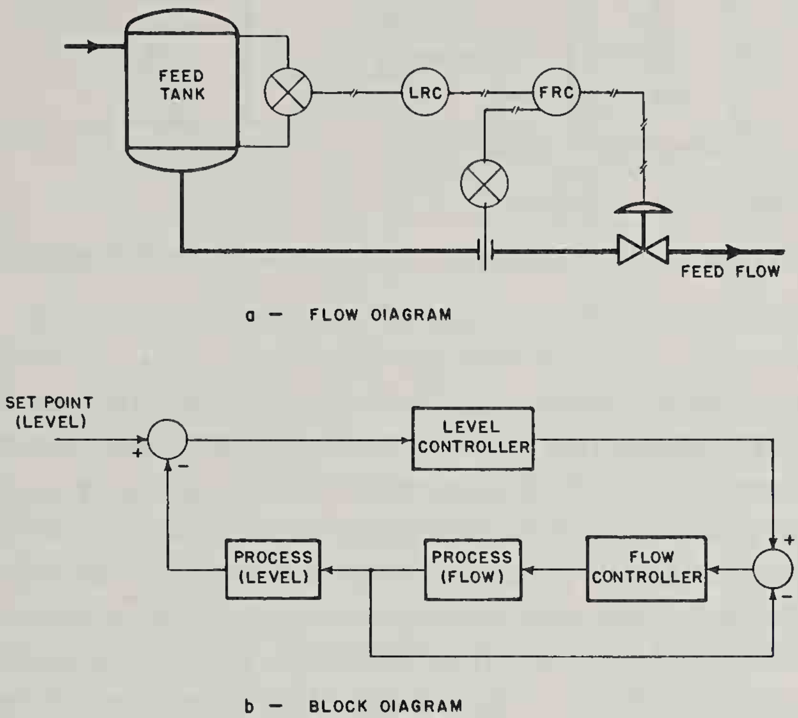


Figure 12–4. Cascade control system.

control loop with a reference input obtained from the level controller. Thus the flow is kept automatically constant at a given set point. Changes of upstream or downstream pressure cannot affect it. As the level changes, however, the reference input signal changes also. This readjusts the set point and with it the flow rate controlled by the flow controller.

It is interesting to observe from the block diagram that this is equivalent to breaking the process into two parts—the "flow" process and the "level" process—and tapping a point between these two parts to obtain an additional feedback loop within the control loop.

Figure 12–5 represents a polymerizer with a cascaded temperature control system. This particular system was designed by the Bristol Co. The polymerizer kettle is heated with hot water which passes through a jacket around the kettle. The temperature inside the kettle must be kept constant. This temperature depends not only on the load conditions of the kettle but also on changes in steam pressure and quality, as well as on changes in cold water pressures and temperature.

The arrangement chosen is to use one temperature recording controller, TRC–1, to measure the temperature of the product inside the

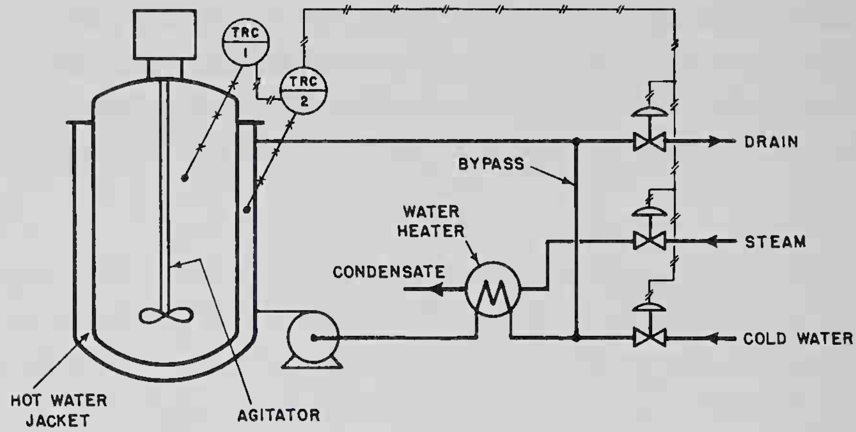


Figure 12–5. Polymerizer with cascade control system.

polymerizer kettle. A second temperature recording controller, TRC–2, measures the temperature in the hot water jacket. This controller is to keep the jacket temperature constant. It accomplishes this by positioning three control valves. As long as the valves in the drain and cold water lines are closed, the hot water circulates through the bypass. These valves are always in identical position and as they open, more and more cold water is admitted, reducing the temperature in the jacket.

The drain and cold water valves are fully open at a signal pressure of 3 psi from controller TRC–2. As the signal increases they close gradually. At 9 psi they are closed completely and remain closed with further increase of pressure. On the other hand, the steam valve remains closed up to 9 psi and begins to open as the pressure rises further, being fully open at 15 psi. Thus at 9 psi all three valves are closed and water circulates but no heat or cold water is added. If the temperature drops below this point, the steam valve begins to open. If the temperature rises, cold water is added.

The TRC–1 controller in this case is a proportional plus rate controller. No reset action is added to avoid overshooting on start up since this is a frequently started batch process. The TRC–2 controller has proportional-position action only.

This process consists of two sharply separated capacitances. One is the capacitance of the jacket, the other is the inside of the kettle. Both capacitances are separated by the inner kettle wall. It has been pointed out previously that the response curve of a multi-capacitance process can be split into a time constant and an equivalent dead time. The disadvantages of dead time, actual or equivalent, have been discussed at length. In the polymerizer control under consideration, the cascade system has actually eliminated this handicap by splitting the capacitances and providing an intermediate feedback signal. Figure 12–6 is the cor-

responding block diagram. In its arrangement, this diagram does not differ from Figure 12–4*b*. The feedback from between the two capacitances, however, illustrates an additional characteristic of feedback, which is an improvement of the dynamic characteristics. Without it, the

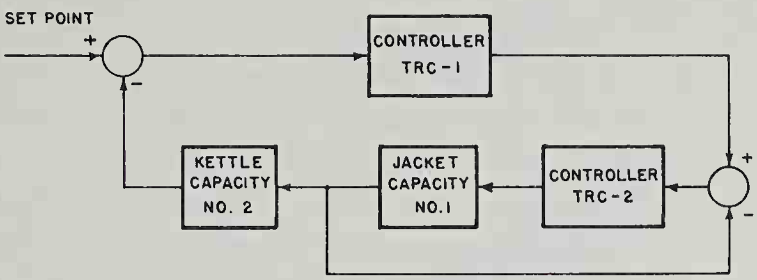


Figure 12–6. Schematic drawing of polymerizer with cascade control system.

detection of temperature changes would be delayed since the heat flow comprises two capacitances with a resistance connecting them. The introduction of a second temperature controller in the jacket signals changes in the process temperature before they actually take place.

The cascade controller requires two or more controllers, each of which contains a complete measuring system and a complete controlling unit. One of these controllers is the master controller. It positions the set point in one or more other controllers, such as secondary, pneumatic set, or autoset controllers. In the block diagram of Figure 12–6 controller TRC–1 is the master controller. The other controller, TRC–2, is the secondary controller.

While the master controller is a conventional one-, two-, or three-term controller, the secondary controller requires modifications. The signal from the master to the secondary controller covers a certain range. In simple cascade controllers, this signal range is equivalent to a change of the set point index in the secondary controller moving from one extreme to the other. This results in certain limitations in adjusting the control actions of both controllers. Manual adjustments are therefore frequently provided to change range and zero of motion of the set point index in the secondary controller.

The secondary controller is generally designed to also permit reversibility of action. There are two interrelated controlled variables in a cascade control system: the master or primary variable and the secondary variable. If an increase of the magnitude of the master variable raises the set point of the secondary variable, the action is called *direct*. If it lowers the set point, it is *reverse action.*

Figure 12–7 illustrates the operation of the Taylor pneumatic set controller. The set point is adjusted pneumatically by the input air sig-

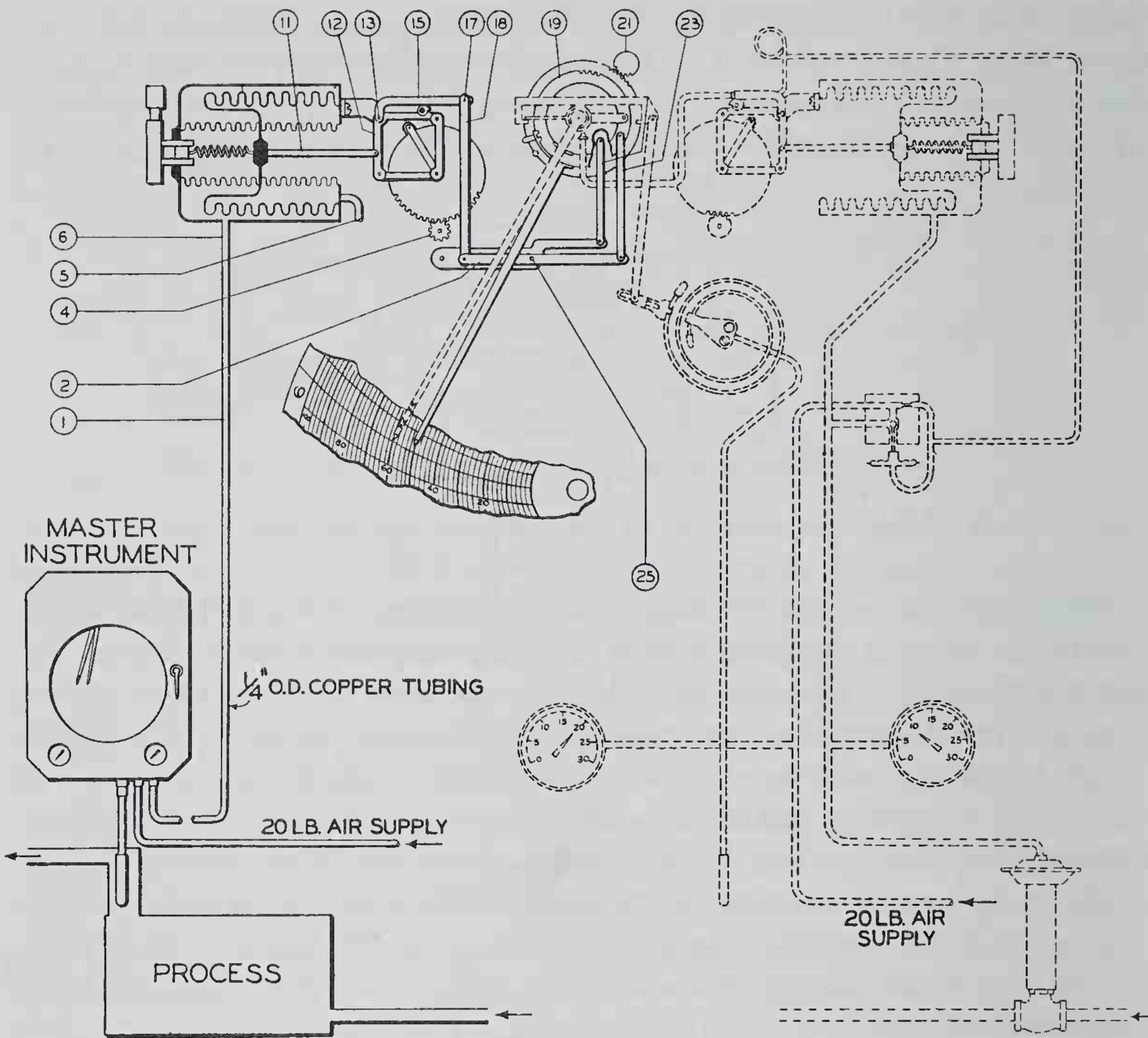


Figure 12–7. Schematic drawing of pneumatic set controller. (*Courtesy of Taylor Instrument Cos.*)

nal from the master instrument. The pneumatic adjustment mechanism is indicated by solid lines to differentiate it from the controller mechanism. With an increase in input air signal entering the pneumatic adjustment unit through line (1) at (6), the set point will be raised. This means that this is direct action. If reverse action is required, line (1) attaches at (5).

With the direct action as shown, an increase in signal moves pin (11) toward the right. Parallelogram (12) transmits this movement to arm (15) which is pivoted at (13), moving it upward. The resulting upward motion of pivot (17) is transmitted through link (18) to bell crank (2) which pivots at (25). Link (23) which attaches to the floating gear (19), to which the set point index is attached, is drawn downward. This moves the set point index to the left.

The amount of travel of the set point index per psi change in input air signal, is regulated by the span adjustment knob (4). Travel increases as the knob is turned to higher dial readings. The dial is calibrated in inches of index travel per psi change in input air signal. It covers a range of 0 to 0.4 in. per psi.

The span adjustment for a required set point index travel per unit change in input air signal can be determined by the use of the equation $A/B = C$, where A is the linear distance of index travel in in., measured on a radius from the center of the chart, B is the total change in input air signal, and C is the setting of the span adjustment knob in in./psi.

For example, if the set point index is to move 3 in. for a variation in input signal from 3 to 15 psi, then the dial is set for

$$\frac{3}{15 - 3} = 0.25 \text{ in. per psi}$$

A span adjustment of 0 in. per psi reduces the pneumatic set controller to a conventional pneumatic controller. It no longer responds to changes in input air signal and controls at the set point which is determined by the adjustment of knob (21).

Various methods are used for determining the upper and lower limit through which the set point index of the secondary controller moves. One method is to make the span adjustment zero. The set point index position under this condition defines the index position at any other span adjustment when the input air signal is 9 psi, i.e., at the midpoint between 3 and 15 psi. Another method is a zero adjustment which sets the lower point of the range through which the set point index moves. In either case the range of set point index motion is shifted either upscale or downscale.

Adjustable control-point limit stops are also frequently built into the secondary controller. If the secondary variable has to be maintained within certain limits, the stops can be so positioned that the control point does not exceed these limits.

Ratio Control

The purpose of ratio control is to maintain the relative magnitude between two variables without being concerned about their absolute magnitude. The simplest form is the open-loop arrangement in Figure 12–8. Flow through line A is measured, and supplies the actuating signal for the flow controller. A valve in line B is positioned by the flow controller. This arrangement contains all the uncertainties of an open loop. The pressures in line B are assumed to be constant and the control valve to be linear. This is obviously an idealized condition which cannot be

obtained. However, in some cases it may be sufficiently approached to be acceptable when only an approximate ratio is required. Under such conditions the example in Figure 12–8 provides a simple and inexpensive arrangement.

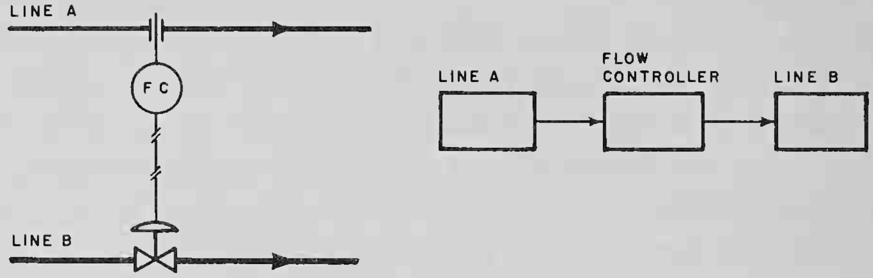


Figure 12–8. Open-loop ratio control system.

Considerable improvement is obtained by the closed-loop method illustrated in Figure 12–9. Flow in line *A*, which is called the primary flow, is now compared with flow in line *B*, called the secondary flow. The controller maintains the secondary flow at a rate which can only be varied when the primary flow changes.

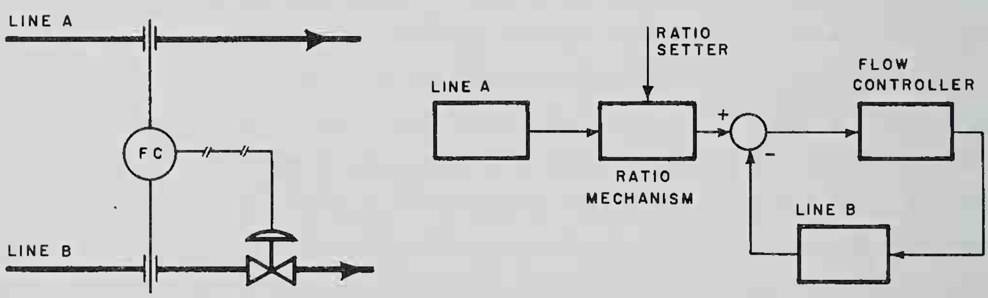


Figure 12–9. Closed loop ratio control system.

The measurement signal from the primary flow r goes into a ratio mechanism. This multiplies the input r by a factor k which is given by the manually adjustable ratio setter. The summation point receives thus a reference input signal equal to kr and combines it with the measurement signal b from the secondary flow. The difference between kr and b is the actuating signal for the controller. The output signal from the controller regulates the secondary flow to make the difference between kr and b equal to zero. Hence

$$kr - b = 0 \quad \text{or} \quad r/b = k \tag{12–4}$$

which means that the ratio between r and b, i.e., between primary and secondary flow, is equal to k. In other words, the control system maintains a ratio which is given by the adjustment of the ratio setter.

Figure 12–10 shows a hydraulic ratio controller. Differential pressure is measured across an orifice plate through which the primary flow passes. This differential pressure is transmitted to diaphragm (1). The force on the diaphragm is applied to counterlever (2) and over the ratio slider (3) to the jet pipe.

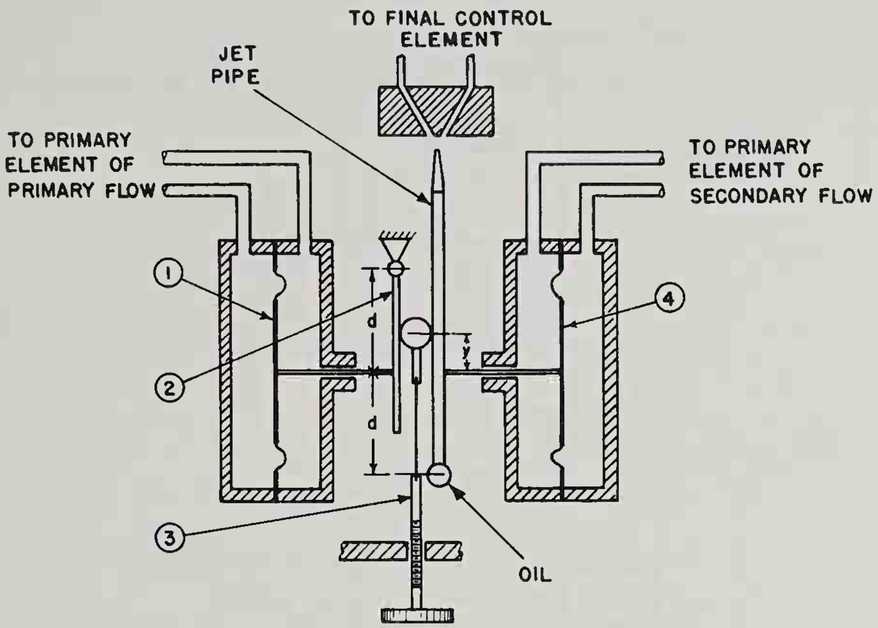


Figure 12–10. Hydraulic ratio control system.

The secondary flow is measured by similar means. The resulting force from diaphragm (4) is directly transmitted to the jet pipe. The position of the jet pipe is a function of the two forces which are due to the primary and secondary flow respectively. These forces balance around the pivot point of the ratio slider, and the equation

$$F_1 \frac{d}{d-y} = F_2 \frac{d}{d+y}$$

is obtained from the lever dimensions d and y as shown in Figure 12–10, using F_1 for the force from the primary flow and F_2 for the force from the secondary flow. Since this shows that

$$\frac{F_1}{F_2} = \frac{d-y}{d+y}$$

it illustrates that the ratio F_1/F_2 can be changed. The dimension d is fixed but dimension y is adjustable by means of the ratio setter.

When either of the two flows changes, the balance between the two forces is disturbed. The jet pipe deflects, thereby providing a controller

output signal for the final control element. The resulting adjustment of the secondary flow reestablishes the ratio balance.

Frequently, the ratio is automatically adjusted in response to a third process variable. This case is illustrated in Figure 12–11 where the feed

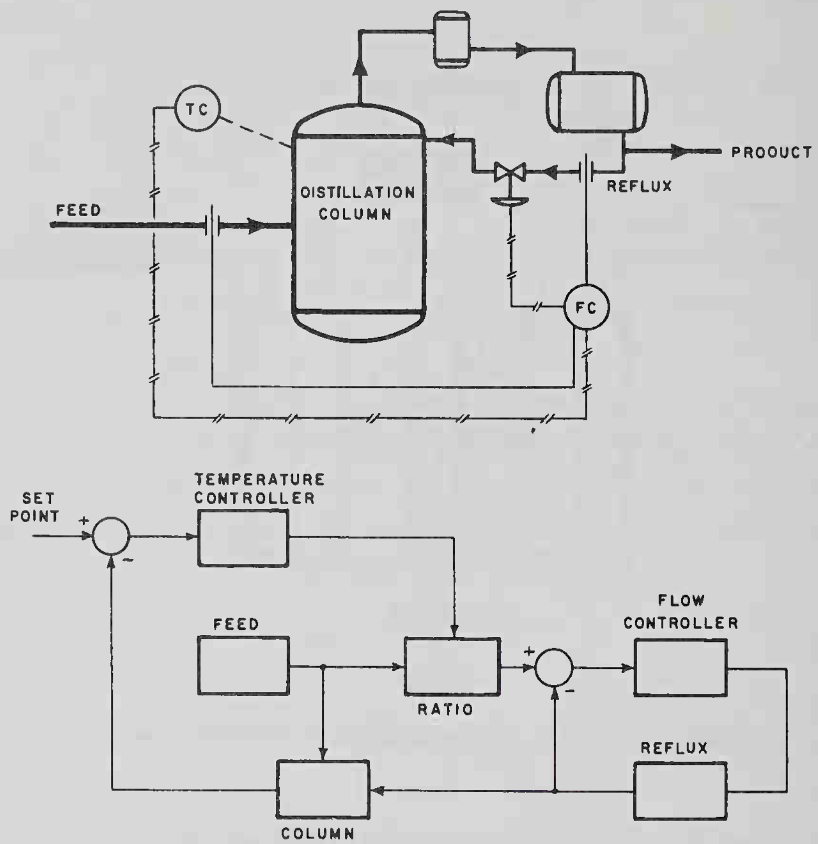


Figure 12–11. Cascade ratio control system.

flow and the reflux flow of a distillation column are maintained at constant ratio. As the temperature rises the ratio of feed to reflux is supposed to increase. As shown, the temperature is measured and the output signal of the temperature controller is utilized to adjust the ratio between feed and reflux. Actually there are two controlled variables, the temperature and the ratio. There are also two complete controllers. Hence this is a cascade system. However, the temperature controller cascades into a ratio system. The block diagram, by showing that one process variable—the feed—is not part of the closed loop, clearly expresses that a non-cascade system is also involved.

2– and 3–Element Control

Equation (12–4) in its first version, i.e., $kr - b = 0$, shows that the ratio controller may be considered a subtracting control system just as well as a ratio system. In fact, the difference is in the viewpoint rather than in

the action. The change in ratio can be considered as altering the set point by subtracting from it some other value. An illustrative example is found in maintaining the level of boiler drums. When the steam demand increases, a momentary "swelling" of the water in the drum occurs. Similarly, when the demand drops, momentary "shrinking" takes place. Response of the level controller to these phenomena produces undesirable upsets of a balanced feedwater flow.

In the two-element system shown in Figure 12–12, the output signal of a flow controller is used to adjust the set point of a level-recording

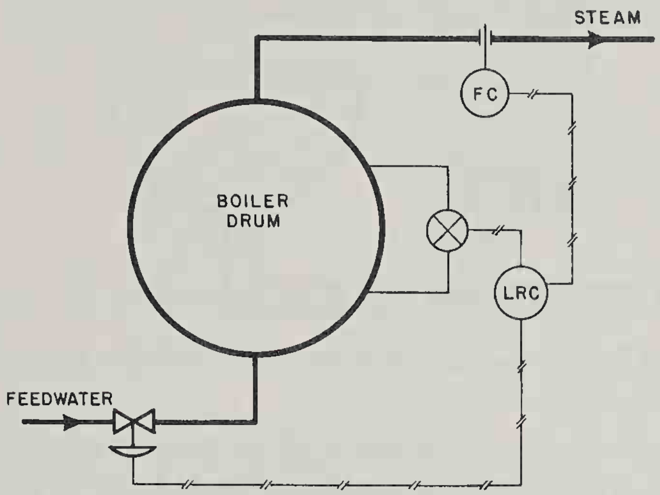


Figure 12–12. Two-element control system.

controller. The level controller regulates the valve position in the feedwater line. As long as the steam flow is constant, the system responds only to changes in level. If the steam flow changes, the set point of the level controller is automatically readjusted to provide for a corresponding change in feedwater flow.

The result is that in case of increased steam flow, the level controller will call for less feedwater flow because of the swelling, while the steam flow controller calls for increase of feedwater flow. The two signals cancel out. The feedwater flow remains constant until the momentary swelling subsides. At that point, the increased demand of the steam flow results in increase of feedwater flow. By this method load changes result in a minimum of upset in the boiler steaming condition.

When a permanent change of set point, and hence of level, cannot be tolerated, a feedback signal from the final control element may balance the flow signal, as well as the level signal. An arrangement of this sort is illustrated in Figure 12–13. Change in steam flow changes the differential pressure across the orifice. The differential pressure is applied across a diaphragm and acts as a force on the beam which positions a four-way

valve. The resulting displacement of the actuating piston operates the control valve. The primary feedback from the level operates in similar fashion. For example, when the level rises, the float motion relaxes the spring that connects it to the lever. This results in a clockwise motion of

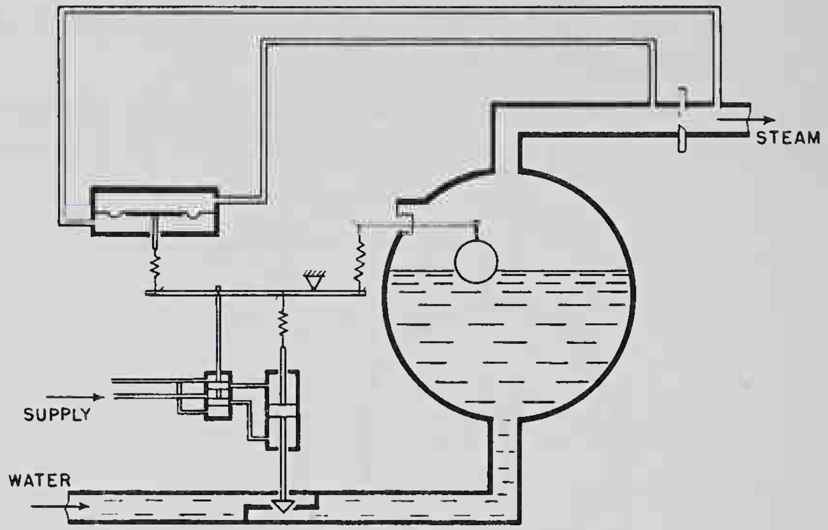
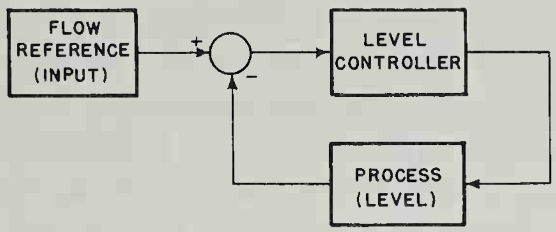


Figure 12–13. Two-element control system with feedback.

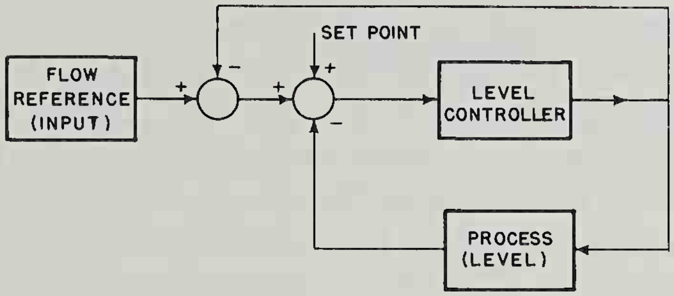
the lever. The four-way valve admits oil to the top of the piston, tending to close the valve. This downward motion feeds back to the level by means of the spring restoring the balance at a new position of the control valve. In case of increased steam demand and momentary swelling, the level tries to move the lever clockwise, while the increased pressure differential opposes this motion. The result is that the control valve does not move until the swelling subsides. Normally, a manual set point adjustment is added, which, however, is not shown in the diagram.

Figure 12–14 illustrates both methods of two-element control. In either case the flow signal is used as reference input for the level control system. The valve-position feedback in the second case, however, subtracts from the flow reference input, and under balanced conditions, the input into the second summation point becomes zero, as far as the flow signal is concerned. The actuating signal into the level controller is then merely the difference between manual set point and primary feedback.

The three-element feedwater control system combines a typical cascade level control system with the steam flow correction. It is illustrated in Figure 12–15 and is particularly suitable where load disturbances in the feedwater supply side are to be expected. The three-element system permits, furthermore, the use of floating control. Since the boiler drum has no self-regulation, floating control cannot be used with it unless a three-element system is used. The latter converts the arrangements into



a — WITHOUT VALVE-POSITION FEEDBACK



b — WITH VALVE-POSITION FEEDBACK

Figure 12–14. Schematic drawings of two-element systems.

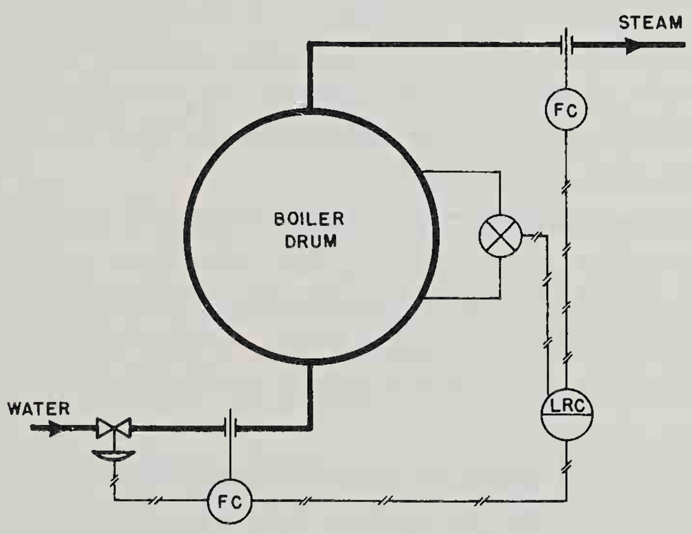


Figure 12–15. Three-element control system.

a flow control system with self-regulation into which the output of the level controller is cascaded.

Totalizing

In oil refineries, steel mills, etc., by-product fuels are frequently burnt either separately or in combination with coal, oil, or natural gas in order to obtain an economical source of thermal energy for steam boilers.

Multiple-fuel combustion presents some interesting control problems. One is that the total Btu supply in fuel must be maintained. Generally,

the by-product is used as base fuel. However, the by-product supply is rarely constant. One, or sometimes two, make-up fuels are therefore required in order to maintain a constant Btu supply for the furnace.

Another problem is the ratioing of air in proportion to the fuel. Frequently, some indirect methods are used, such as maintaining a constant ratio between steam and air flows in a steam boiler. Flue gas analysis may also be employed in order to proportion air to the oxygen content of the flue gas.

The most direct method, however, is the direct ratioing of air to fuel, either separately for each fuel or for the sum total of the fuels. As compared with gas analysis methods, this provides faster control, since correction can be made at the moment when the fuel flow rate changes, and not after the results of the change become measurable in the flue gases.

A disadvantage, however, of proportioning air to fuel flow is a resultant volumetric ratio. The correct ratio between air and fuel is based on weight. If either the air or a gaseous fuel change in pressure or temperature, the ratio will lose accuracy, unless provisions are made which correct for these changes.

Figure 12–16 illustrates the principle of the Hays fuel totalizing control system. Total Btu input is the controlled variable of this system. The desired magnitude of total Btu is adjusted by means of the Btu control potentiometer either manually or automatically by a steam pressure controller or temperature controller.

In the arrangement shown, either fuel 1 or fuel 2 may be the base fuel, and the remaining fuel, the make-up fuel. The flow rate of the base fuel may vary, yet the total Btu input is maintained at the desired value by automatic adjustment of the make-up fuel. Which one of the two fuels is to be the base fuel and which the make-up fuel can be determined by a manual selector switch.

Flow meters measure the flow rates of the two fuels. These meters are provided with mechanisms for square root extraction, so that the output is proportional to flow rate. A potentiometer in each flow meter is positioned in proportion to the rate of the respective flow.

The output of the two potentiometers is averaged by use of a center-tapped autotransformer, and this average voltage is applied to the coils of the relay. The relay contacts operate a valve actuator which adjusts the control valve of the make-up fuel. When the relay closes one contact, the actuator opens the valve; when it closes the other contact, the valve closes. The valve maintains its position when none of the contacts is closed.

With the Btu control potentiometer in a given position representing a definite Btu input demand, and the base fuel being burned in a quantity

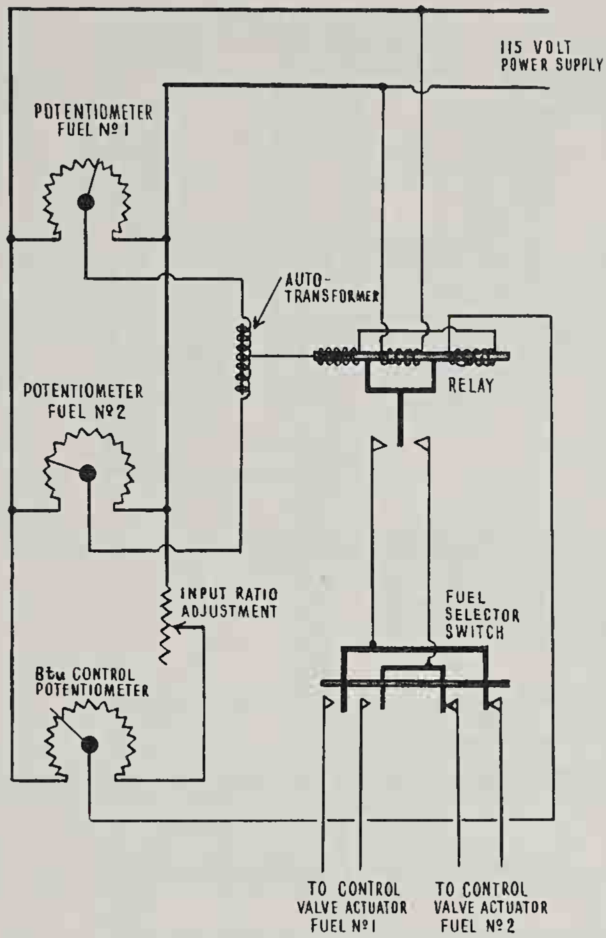


Figure 12–16. Principle of Hays fuel totalizing control system.

less than sufficient to meet the demand, the electrical relay closes the corresponding contact. This opens the valve of the base fuel until the flow as measured by the flow meter reaches the desired level, and the relay contact is again opened.

Figure 12–17 shows the principle of an Askania multiple fuel control system. In this case, two fuels are burnt in a furnace and air is supplied to match the fuels. The arrangement is as follows:

A small blower or compressed air source provides a pilot air flow through two lines, each about one inch in diameter. A ratio controller *A* measures on its primary diaphragm the differential pressure across an orifice plate in the fuel line I. By means of a hydraulic flow through a jet pipe, this controller positions the actuating piston of the control valve in *C* of the pilot air flow arrangement. An orifice plate in the *C* branch provides a differential pressure that is imposed on the secondary diaphragm of the controller *A*. A rate of pilot air flow is thus obtained in *C*, which is directly proportional to the rate of flow of fuel I.

Ratio controller *B* in a like manner measures the flow of fuel by means of an orifice plate in fuel line II, and in turn provides a means of con-

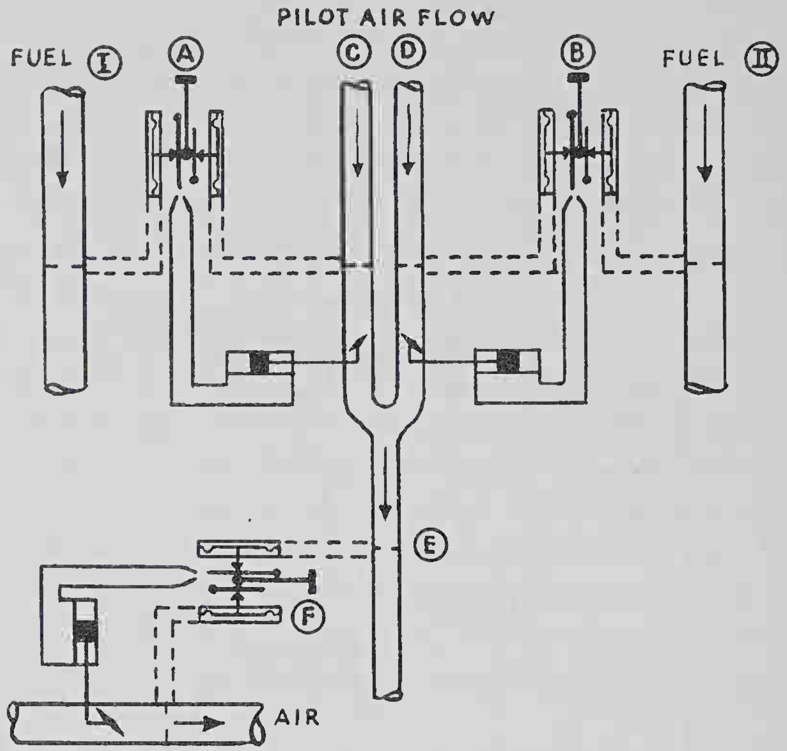


Figure 12–17. Askania totalizer and air ratio control system.

trolling an air pilot flow in proportion to this fuel, and thus to its air requirement.

Therefore, controllers *A* and *B* regulate the rate of air flows *C* and *D* in such a way that the air flow in each line is always proportional to the flow of each fuel. These air flows are then brought together and the total flow, measured by means of orifice *E*, provides the primary signal for ratio controller *F*. The combustion air is regulated by this last controller, which maintains it in fixed proportion to the common pilot air flow which represents the sums of fuels I and II.

The ratio sliders *A* and *B* may be set for different air ratios corresponding to the requirements of each fuel. They can also be automatically set as a function of temperature and pressure of gaseous fuels, and thus correct for changes in density. The same holds true of the combustion air ratio controller.

Time-Schedule Control

Time-schedule controllers regulate batch processes according to some predetermined time schedule. Various methods are described here and illustrated in Figure 12–18; these are available in the line of time-schedule controllers made by the Foxboro Co.

The *automatic shutdown controller* starts its cycle when a push button is depressed. This opens the valve and leaves it open until the controlled

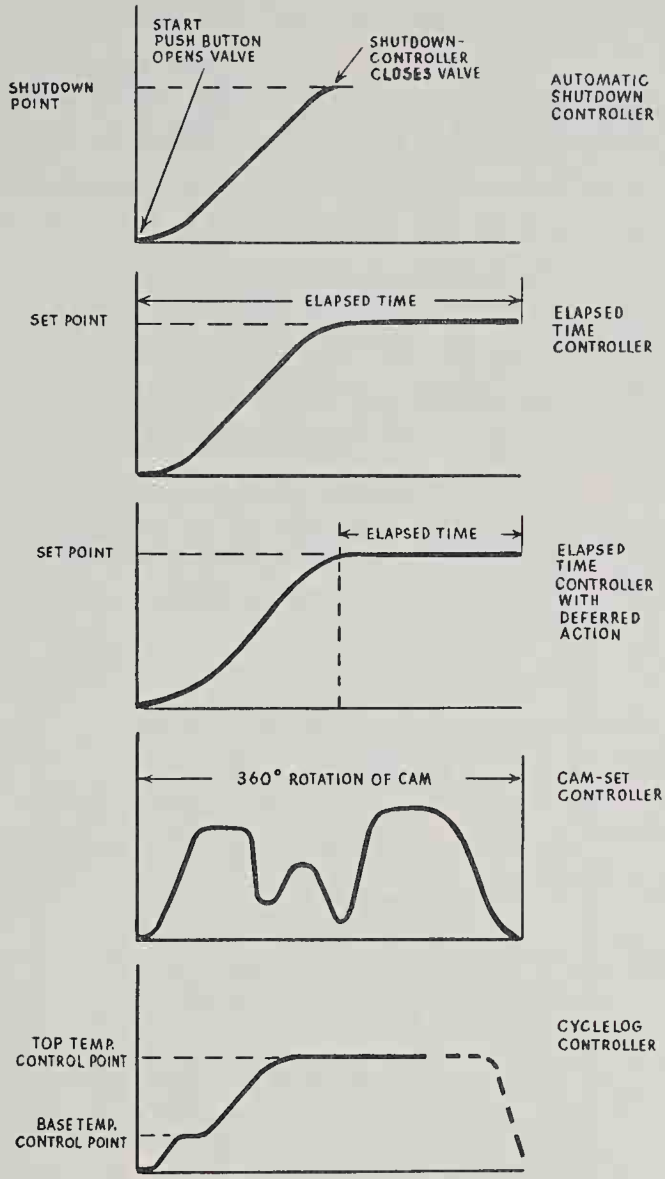


Figure 12–18. Action from various types of Foxboro time-schedule controllers.

variable reaches a predetermined value, the shutdown point. The valve then shuts down automatically.

The *elapsed time controller* starts a preselected time period. Within this time, it brings the controlled variable to its set point and then controls it at that point. At the end of this period the valve is automatically closed.

The *elapsed time controller with deferred action* differs in that the elapsed time is counted from the time where the controlled variable reaches the set point.

The *cam-set controller* regulates a schedule according to the contour of a rotating cam. Where the full 360 degree cycle of a rotating cam is not

required, the cam is returned to the starting point by a manual adjustment.

The *Cyclelog controller* is used to pre-heat a batch to any selected base temperature. It then holds the controlled variable at that temperature for a preset period of time. After this time the temperature is allowed to rise at a preselected rate to the top set point. It holds the batch at the top set point for a certain period of time, after which it automatically shuts down the process. Two steps may be added as indicated by the dashed line in Figure 12–18. They consist of (1) a predetermined prolongation of the time which the valve stays open by pressing a push button, and (2) the provision of a controlled rate of cooling.

These are only a few of a large number of variations which are used in control systems where the control of a process becomes a time function.

13. INDUSTRIAL APPLICATIONS

Only experience can teach the proper application of the principles of automatic controls. New combinations of automatic controllers and their accessories come up continuously. Some practical examples, taken from a number of industries, are described in the following pages.

Steam Generation

Figure 13–1 shows the automatic control system of a steam-generating apparatus. It follows closely an installation by the Hagan Corporation

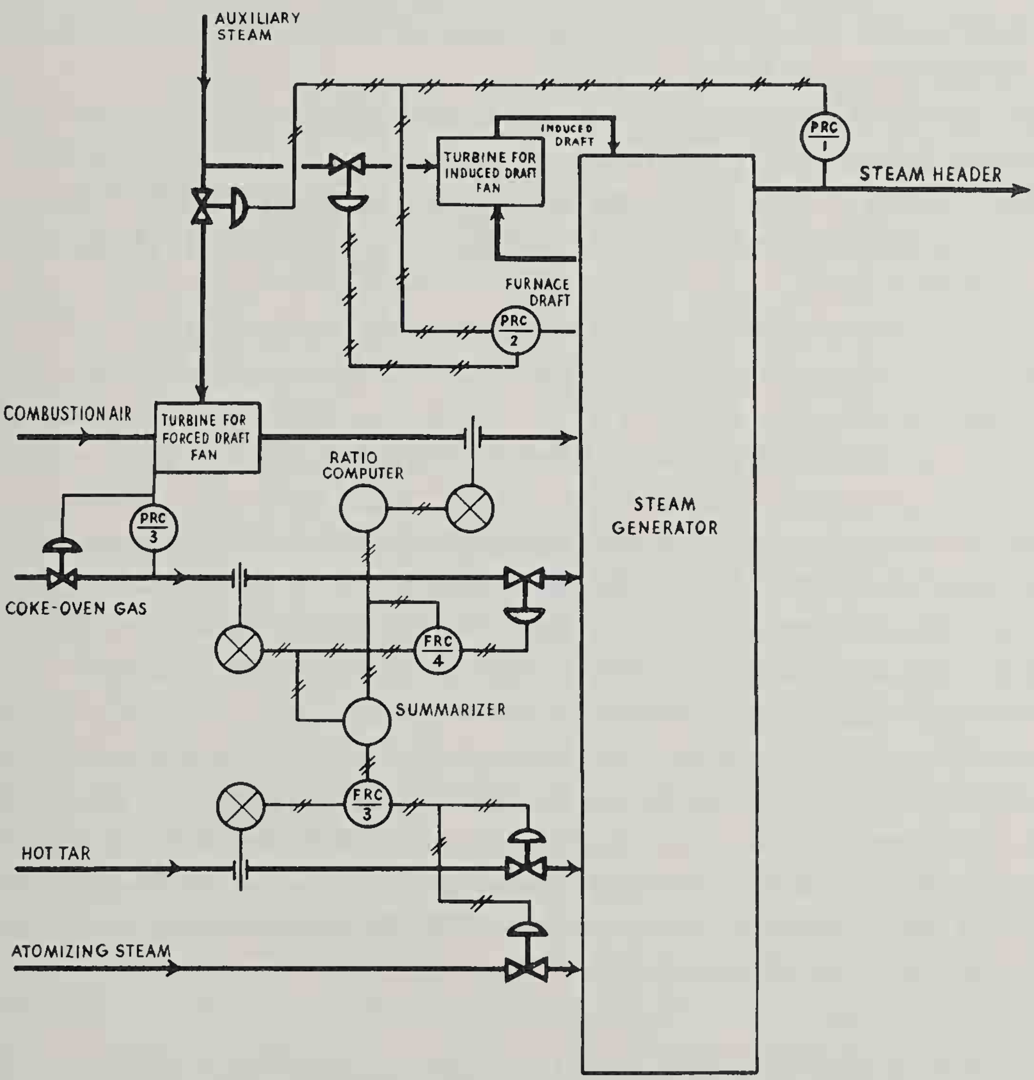


Figure 13–1. Automatic control system of steam generator.

at the Hazelwood By-Product Plant of Jones & Laughlin Steel Corporation.* Coke-oven gas is used as base fuel and tar is used to make up for deficiencies in the supply of coke-oven gas. The tar is preheated to a temperature of 250°F which makes its consistency comparable to that of fuel oil. Atomizing steam is added to the tar for easy combustion.

The pressure in the steam header is a direct function of the demand on the steam generator. This pressure controller is the master controller of the entire combustion system. The sequence by which it operates is as follows.

Suppose the steam header pressure drops. This is sensed by pressure controller PRC–1 and results in an increase of auxiliary steam to the turbine which drives the forced draft fan. The increased rate of flow of combustion air is measured and changes the set point of flow ratio controller FRC–4. A ratio computer converts the air flow signal into a proportional signal which is applied to the ratio controller. This raises the rate of flow of coke-oven gas. The increased input of fuel in proportion to combustion air results in an increase of steam generation, returning the steam pressure to its set point.

The pressure of the coke-oven gas is maintained by controller PRC–3. This has the advantage of constant upstream pressure for the flow controller, which is thus independent of varying pressure drops in the gas lines resulting from changes in flow rate due to position changes in the flow control valve.

The coke-oven gas flow is maintained constant by the FRC–4 flow controller, but it changes in proportion to the combustion air flow. A summarizer provides a signal to flow ratio controller FRC–3. This signal is equal to the fuel demand from the ratio computer minus the fuel supply as measured in the coke-oven gas line. As long as the coke-oven gas copes with the fuel demand, the control valves in the tar and atomizing steam lines are closed. When a deficiency is signaled by the summarizer, the set point of the FRC–3 controller is automatically raised and hot tar and atomizing steam start to flow simultaneously.

Furnace draft is controlled to limit the amount of cold air filtering from the outside through the brickwork of the furnace and to protect from the danger of back firing in case of over-pressure in the combustion chamber. The furnace draft controller PRC–2 is a ratio controller. Steam pressure is the primary variable. The final control element is a valve in the auxiliary steam element which controls the steam to the turbine which drives the induced draft fan.

---

* Peth, H. W., Kotsch, J. A., and Gilmer, H., "Highly Automatic Steam Generation," *Instruments and Automation,* (Dec. 1955).

In addition to the control system illustrated in Figure 13–1, the flame in the combustion chamber is continuously monitored. This prevents the flow of coke-oven gas or tar into the furnace without immediate combustion. The combustion detector is connected into an electric relay system which shuts down air, gas, and tar flow in case of flame failure. This is done by additional valves not shown in the diagram. The same electric relay system is utilized to provide shutdown in case of fan failure, either in the induced draft or the forced draft system, of feedwater failure, or of power failure.

Soaking Pit

Before steel ingots pass to the rolling mill where they are pressed to their ultimate shape—sheet metal, bars, etc.—they are heated in a soaking pit to a high, uniform temperature. The most desirable condition is a uniform plasticity at the highest possible temperature without melting any of its constituent elements. Figure 13–2 shows a control system used

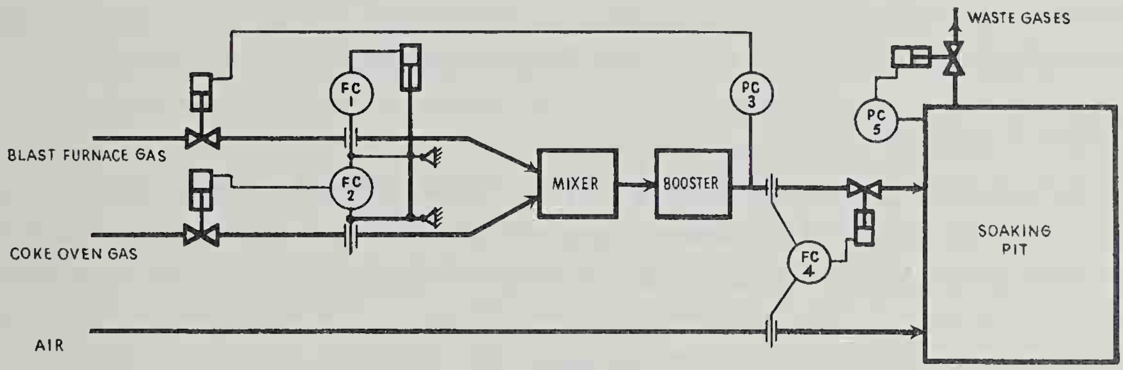


Figure 13–2. Automatic control system of soaking pit.

by Askania Regulator Co. for premixed gas-fired soaking pits. Blast furnace gas and coke-oven gas are used as fuel. They are combined in a mixer and pass from there to a booster where they are brought to the pressure required by the burners. This pressure is measured by pressure controller PC–3, and is controlled by regulating a butterfly valve in the blast furnace gas line supplying the mixer. For every change in gas load there will be a corresponding change in the blast furnace gas supply in order to maintain the mixed gas pressure at the desired set point.

The primary elements which measure the rates of flow of blast furnace gas and coke-oven gas are butterfly valves which substitute the conventional orifice plate. The differential across the butterfly valve is measured as a function of the flow rate. Flow controller FC–1 changes the opening of the butterfly valves with changes in flow. Hence the equivalent of a variable area meter is obtained. This provides greater accuracy through-

out the control range and avoids the limitations in rangeability that are characteristic for fixed orifice plates.

Proportioning of the blast furnace gas and coke-oven gas is provided by flow ratio controller FC–2. The primary variable is the blast furnace gas. When the rate of flow of the blast furnace gas changes, the set point of FC–2 is automatically readjusted and the control valve in the coke-oven gas line is repositioned to maintain the constant ratio between the two gases. Another flow ratio controller, FC–4, proportions fuel gas to air flow.

PC–5 is the soaking pit pressure controller. The pressure is measured in one of the side walls and transmitted to the controller which positions a damper in the stack through which the waste gases escape. Positioning the damper controls the pressure in the soaking pit. This prevents suction in the pit which may otherwise result at low fuel rates and produce excessive air infiltration. Not only is heat loss the result of air infiltration, but the flow of gases in the pit itself may be disturbed, upsetting the uniformity of heating in the various sections. On the other hand, if excessive pressure builds up, heat is wasted at openings, cracks, and through the soaking pit walls, resulting in serious economic losses.

Water Treatment

The automatic control system of a water treatment plant is illustrated in Figure 13–3. The diagram is based on the Honeywell control system* at McAlester, Oklahoma. Water is pumped from a large lake reservoir

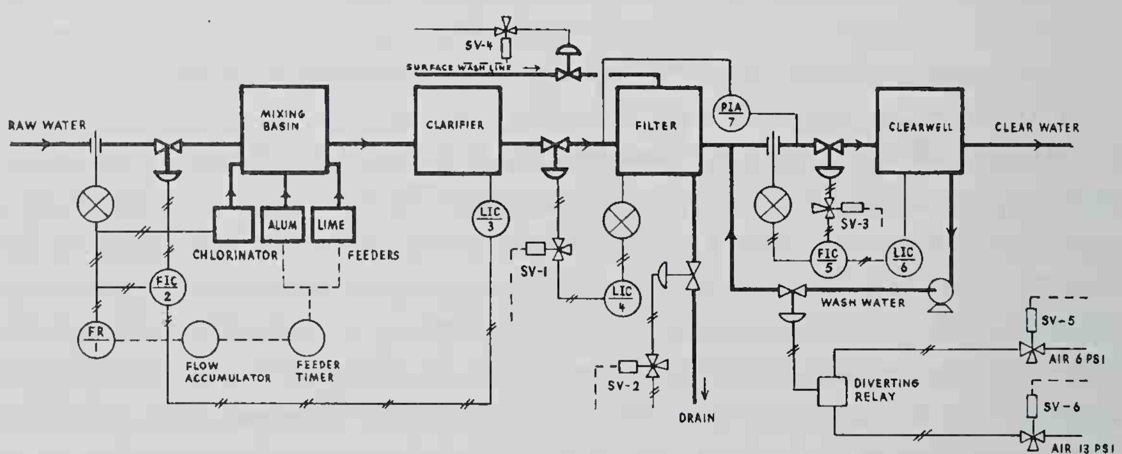


Figure 13–3. Automatic control system of water treatment plant.

to a second lake for storage. From there it flows by gravity toward the mixing basin, shown in the illustration.

* Collins, R., "McAlester Modernizes," *Instrumentation*, Vol. 10, No. 2, (1957).

The flow into the mixing basin is controlled by the flow ratio controller FIC–2. When the level in the clarifier drops, the level controller LIC–3 adjusts automatically the set point in FIC–2 to admit more flow to the treatment plant.

The measured flow is signalled to the chlorinator which adds chlorine to the water in proportion to rate of flow. Furthermore, flow meter FR–1 records and integrates the raw water flow. A flow accumulator which parallels the flow meter integrator starts the feeder timer after each 2315 gallons of raw water. The feeder timer actuates both alum and lime feeders and adds these chemicals to the raw water for 30 seconds. This process reduces the turbidity of the water to less than 10 parts per million.

The water then flows by gravity into a clarifier, then into sand filters and on to a million-gallon clear well. No measurable concentration of suspended solids are contained in the clear well water. The chlorine residual is about 0.4 parts per million.

Flow to the clear well is regulated by ratio flow controller FIC–5 with level controller LIC–6 for automatic set point adjustment.

The pressure drop across the filter is indicated by PIA–7. As the filters get clogged, the pressure drop increases. It is allowed to rise a certain amount, but when this amount is exceeded, an alarm sounds. This signals that the filter needs cleaning. The operator starts the backwashing cycle, which removes the filter from the continuous water supply cycle. Since, however, a number of filters (not shown in the diagram) are parallel, the load is taken over by them and the process is not interrupted.

When the alarm sounds, the operator presses two buttons: one to silence and reset the alarm, the other to start a timer which rotates eight cams on a common shaft, making and breaking certain electrical contacts to produce the backwashing cycle as follows:

Solenoid valve SV–1 is energized, shutting off the controller output signal from LIC–4 and releasing the air to the control valve. This closes the filter influent valve.

After ten minutes, the timer energizes solenoid valve SV–2, admitting air to the drain valve and opening it. At the same time SV–3 is also energized, which closes the filter effluent valve. Now the solenoid valve SV–4 is energized and opens the valve in the surface wash line for one minute. This done, solenoid valve SV–5 admits air of 6 psi pressure through the diverting relay to the valve in the wash water line. The diverting relay has the characteristic of always passing the air which has the highest pressure. After 2.5 min. solenoid valve SV–6 energizes and 13 psi air pressure now flows to the control valve. This sequence of a lower and then a higher pressure produces first a low baskwash rate followed

by a faster backwash rate. After another five minutes SV–6 is deenergized and the system returns to a low backwash rate for 2.5 min. Remaining steps return the system to normal operation by closing the wash and drain valves and opening the influent and effluent valves. The effluent valve is opened only after the filter has been refilled to proper height.

Air Conditioning

The control of humidity and temperature in an air conditioning system is illustrated in Figure 13–4. The air which flows into the humidifier becomes saturated after being sprayed with water. This water is main-

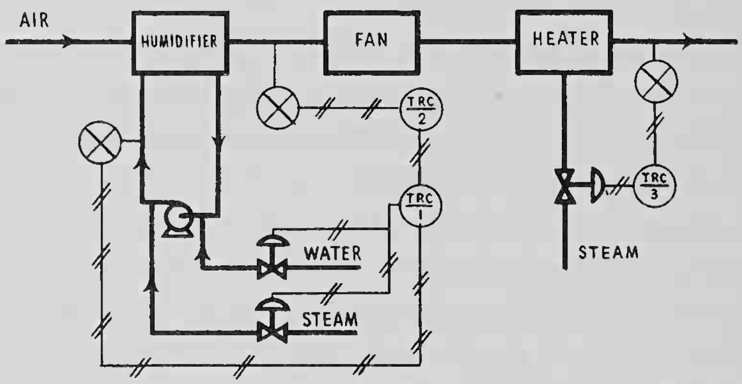


Figure 13–4. Air conditioning control system.

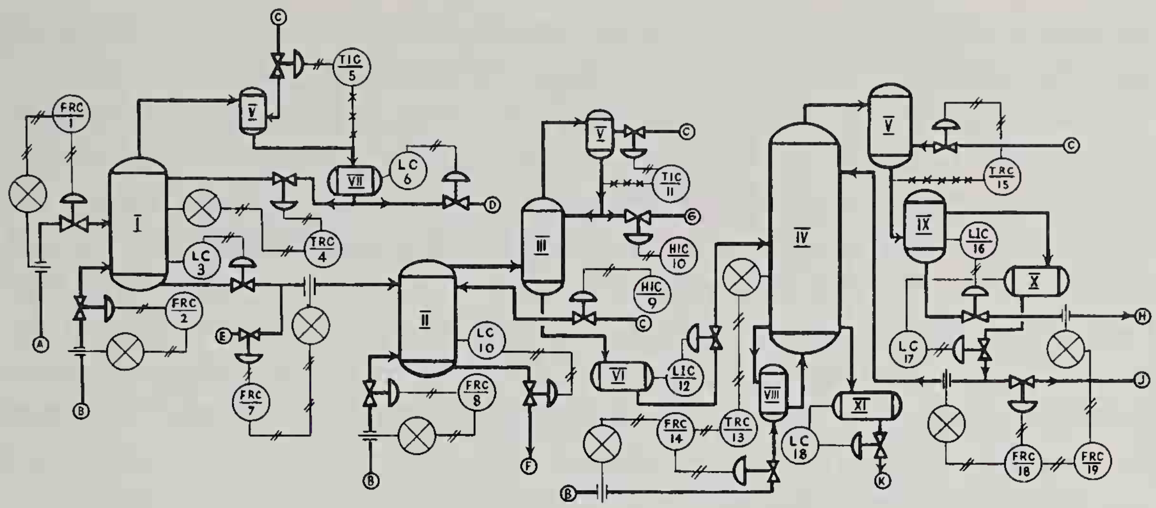
tained at constant temperature by means of split range valves for water, and is steam controlled from a cascade temperature controller. The humidifier represents a process capacitance, and since the response of TRC–2 is delayed more than is suitable for good control, it is aided by controller TRC–1 to improve the dynamic response.

The saturated air then passes through a heater where, by means of steam coils, the temperature is brought to that maintained by temperature controller TRC–3.

Solvent Recovery

Figure 13–5 illustrates the control system of a solvent recovery unit. The diagram is based on an installation* by Taylor Instrument Companies at Lederle Laboratories, Pearl River, New York. This unit recovers organic solvents which are needed in large quantities for obtaining aureomycin from fermentation mash. The unit consists of four distillation columns. The incoming feed contains four component solvents which are designated *a, b, c,* and *d.* In the first column, a liquid mixture of *a,*

* Englund, S. W., and Giesse, R. C., "Solvent Recovery at Lederle Labs," *Taylor Technology,* Vol. 5, No. 2, (1952).



A — Feed I with solvents o, b, c ond d
B — Steam
C — Water
D — Product with solvents a ond b
E — Feed II with solvents c, d ond H_2O
F — To woste
G — Solvent b
H — Waste layer for recavery
J — Salvent c
K — Salvent d

I — First Column
II — Secand Calumn
III — Third Calumn
IV — Fourth Calumn
V — Condenser
VI — Feed Drum

VII — Reflux drum
VIII — Colondria
IX — Decanter
X — Reflux drum
XI — Pump tank

Figure 13–5. Automatic control system of solvent recovery unit.

b and water is distilled off. The remaining solution passes on to column II and is combined with a second feed flow containing *c, d,* and water. Solvent *b* is recovered from column III, and solvents *c* and *d* are recovered from the overhead and bottom product, respectively, of column IV. These recovered solvents are in essentially pure form and are returned to the production process. The bottom layer in the decanter is passed on as waste to another column which is not shown in the diagram.

Feed flow and steam are admitted to column I in controlled quantities. The distillate leaves at the top and passes to a condenser where it is liquefied. From there it flows to the reflux drum and is then either returned to the column for another distillation cycle or drained as product containing *a* and *b* components. The amount of reflux is controlled by the recording temperature controller TRC–4. The condensate temperature is also measured and regulates the rate of flow to the condenser.

The flow of bottom product from column I to II is combined with another feed flow. The total flow is maintained constant by FRC–7 admitting as much feed flow as required for maintaining the constant flow rate. Water is admitted to column II by means of a manually controlled valve with a position adjuster at the control panel. The development of suitable analysis control instruments would probably allow automatic

control in this case. As it is, bottom and overhead analyses are made at periodic intervals. The bottom product of column II goes to waste.

Column III sends its overhead through reflux and drains periodically by another manually controlled valve to recover solvent *b*. The bottom product passes through a feed drum with level control into column IV. Heat is supplied by means of a calandria through which part of the liquid circulates. The steam flow is kept constant by ratio controller FRC–14. If the column temperature drops, the rate of steam flow is increased by readjusting the set point of FRC–14 by means of temperature recording controller TRC–13. The bottom product of column IV passes into a pump tank. It consists of solvent *d*. The overhead is condensed and flows from the condenser to a decanter. The bottom layer flows out to another column. The top solvent layer goes to the reflux drum, and is then split into reflux and solvent *c*. The reflux is maintained constant by ratio controller FRC–18. Any excess of this flow is drained as solvent *c*. An inverse ratio between reflux and waste layer flow from the decanter is maintained by readjusting the set point of FRC–18 by means of FRC–19. This is done to return sufficient solvent to the column, though some is removed to the additional column.

Gasoline Production

A control diagram for a Perco motor fuel alkylation process is shown in Figure 13–6. The arrangement is based upon a control system proposed by Berger and Peters,* using a number of analytical instruments for control purposes. This process takes olefins, i.e., propylene, butylenes, and amylenes, and alkylates them with isobutane, using liquid hydrofluoric acid as catalyst. Non-gasoline-range hydrocarbons are thus converted into high-octane blending stocks.

It is not intended to analyze the process in detail but rather give an over-all concept and then show how some of the variables are controlled. An electronic control system with electrohydraulic valve actuators is assumed.

The liquid olefins enter the process at *A*, while isobutane for alkylation is added through *B*. Both feeds, after passing through their respective driers, are combined with additional recycled isobutane, and enter the reactor II. Hydrofluoric acid, the catalyst, enters from the bottom.

After the alkylation reaction is completed in the reactor, the mixture flows into the settler III, where the hydrofluoric acid separates from the hydrocarbons. Most of the acid is recycled to the reactor. A portion,

---

* Berger, D. E., and Peters, W. D., "Converting Non-Gasoline-Range Hydrocarbons into High-Octane Blending Stocks," *Control Engineering*, (Sept., 1956).

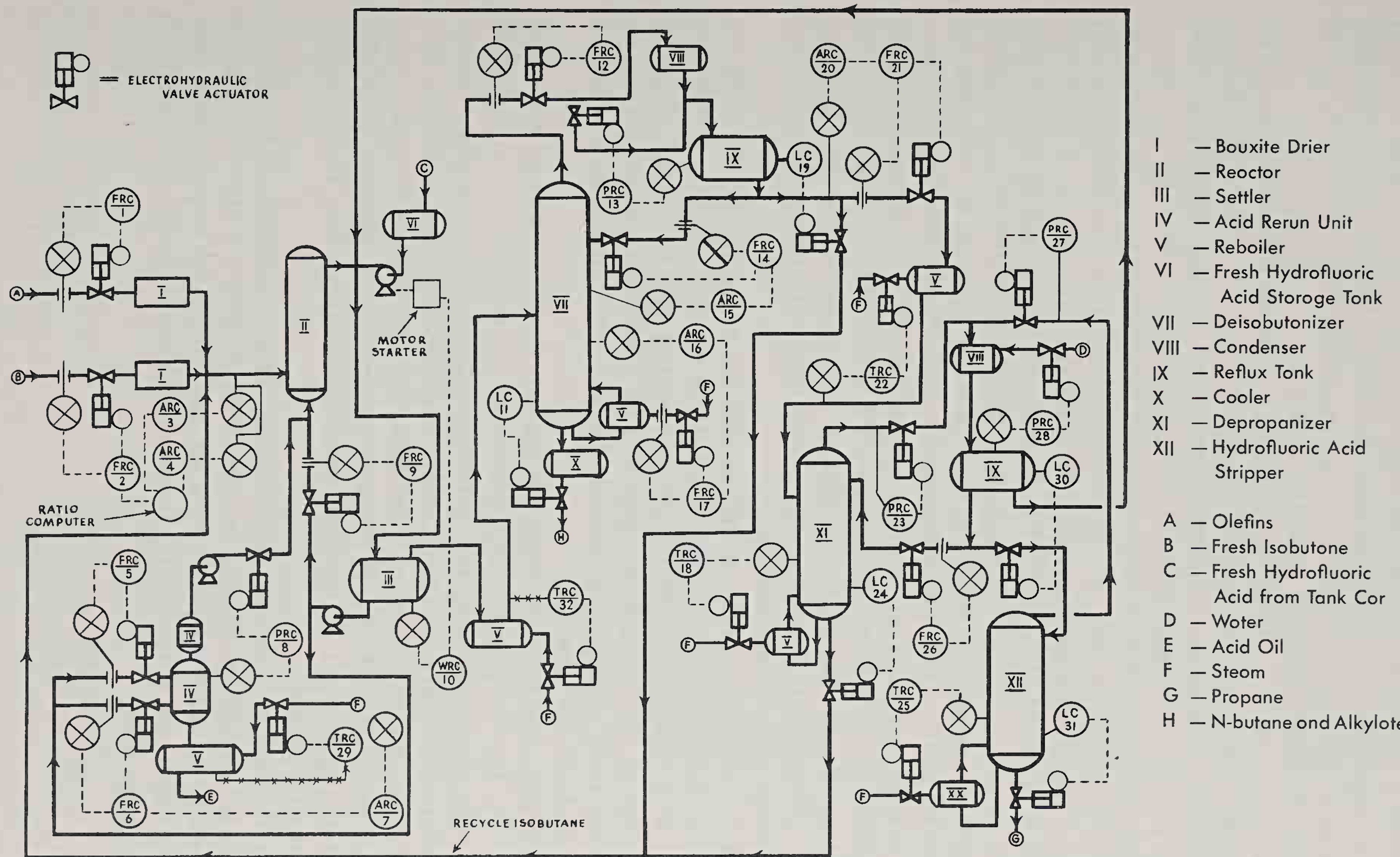


Figure 13–6. Automatic control system for Perco motor fuel alkylation process.

however, is deviated through the acid rerun unit where water and acid oils are separated from the hydrofluoric acid by distillation. To replenish the acid in the system, a fresh acid storage tank, VI, is provided, which receives its charge directly from the tank car.

The hydrocarbons from the settler flow into the deisobutanizer, VII, where the isobutane and lighter components are separated. The bottom product of this column is normal butane and alkylate which is drained off at *H*. The overhead product provides recycle isobutane for the reactor. A small portion of this isobutane, however, is sent for fractionation to the depropanizer, XI, to obtain upgraded recycle isobutane, and to the hydrofluoric acid stripper, XII, for the production of propane, which is removed at *G*. The hydrofluoric acid is recycled to the reactor.

The diagram of this process is equipped with a number of analytical control instruments. Thus, the feed into the reactor is analyzed for olefins by ARC–3 and for isobutane by ARC–4. A ratio computer permits setting the ratio of these two components. The signal from this computer to the flow ratio controller FRC–2 is the difference between the desired and the measured ratios, and readjusts the set point of FRC–2. Since FRC–2 controls the isobutane, it will change this flow until the desired ratio is reestablished.

ARC–7 analyzes the hydrofluoric acid for water and acid-soluble oil contents. It controls the set point of FRC–6 which controls the feed flow to the acid rerun unit. This feed flow is heated by a heater not shown in the diagram. A second flow controlled by FRC–5 by-passes the heater. Thus the ratio of these two flows is determined by the purity of the recycle hydrofluoric acid.

The reflux into the deisobutanizer is controlled by flow ratio controller FRC–14. The set point of this controller is automatically adjusted by ARC–15 which analyzes for N-butane concentration in the column. Another analytical instrument, ARC–16, determines the isobutane concentration and controls the steam input into the reboiler by readjusting the set point of FRC–17. A propane concentration analyzer, ARC–20, controls the amount of deisobutanizer overhead product that is passed into the propanizer.

Another interesting arrangement is the weight controller WRC–10. By means of a strain gauge, the weight of the contents in the settler is determined. When it drops under a set minimum, the two-position controller starts the motor pump to get additional hydrofluoric acid from the storage tank.

INDEX